

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

Цифрова обробка сигналів та зображень

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського як
навчальний посібник для здобувачів ступеня магістра
за освітньою програмою «Комп'ютерно-інтегровані системи та технології в
приладобудуванні» спеціальності 174 (G7) Автоматизація, комп'ютерно-інтегровані
технології та робототехніка

Укладачі: В. І. Микитенко, Г. С. Тимчик

Електронне мережеве навчальне видання

КИЇВ
КПІ ІМ. ІГОРЯ СІКОРСЬКОГО
2025

УДК 621.3
К61

Укладачі: *Микитенко Володимир Іванович*, д-р техн. наук, проф.
Тимчик Григорій Семенович, д-р техн. наук, проф.

Рецензент *Котовський В.Й.*, докт. техн. наук, професор кафедри загальної фізики та моделювання фізичних процесів ФМФ, КПІ ім. Ігоря Сікорського

Відповідальний редактор *Стельмах Н.В.*, канд. техн. наук, доцент

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 3 від 07.12.2024 р.) за поданням вченої ради приладобудівного факультету (протокол № 9/24 від 30.10.2024 р.)

К61 Цифрова обробка сигналів та зображень. [Електронний ресурс] : навч. посіб. для здобувачів ступеня магістра за освіт. програмою «Комп'ютерно-інтегровані системи та технології в приладобудуванні» спец. 174 Автоматизація, комп'ютерно-інтегровані технології та робототехніка / КПІ ім. Ігоря Сікорського ; уклад.: В. І. Микитенко, Г. С. Тимчик. Електрон. текст. дані (1 файл). – Київ : КПІ ім. Ігоря Сікорського, 2025. – **181 с.**

Навчальний посібник «Цифрова обробка сигналів та зображень» містить інформацію про основні типи сигналів, що вимірюються технічними системами, а також про найбільш розповсюджені методи аналізу й обробки сигналів в часовій та просторовій областях. Увагу приділено як аналоговим, так і дискретним сигналам. Призначений для здобувачів ступеня магістра за освітньою програмою «Комп'ютерно-інтегровані системи та технології в приладобудуванні» спеціальності 174 (G7) Автоматизація, комп'ютерно-інтегровані технології та робототехніка. Також буде корисним інженерним працівникам, пов'язаним з вимірюванням сигналів та робототехнікою

УДК 621.3

Реєстр. № НП **XX/XX-XXX**. Обсяг **7,5** авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського» проспект
Берестейський, 37, м. Київ, 03056
<https://kpi.ua>

Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

© КПІ ім. Ігоря Сікорського, 20**25**

Зміст

ВСТУП.....	5
РОЗДІЛ 1. КЛАСИФІКАЦІЯ ТА ОСНОВИ ЦИФРОВОГО ПЕРЕДСТАВЛЕННЯ СИГНАЛІВ.....	7
1.1 Вступ до цифрової обробки сигналів та зображень. Класифікація вимірювальних сигналів.....	7
1.2 Математичне представлення сигналів. Тестові сигнали.....	20
1.3 Узагальнені схеми цифрових систем.....	24
1.4 Аналого-цифрове перетворення сигналів	27
1.5 Цифро-аналогове перетворення сигналів.....	31
Питання для самоперевірки.....	37
РОЗДІЛ 2. СПЕКТРАЛЬНИЙ АНАЛІЗ СИГНАЛІВ. ІНТЕГРАЛЬНІ ПЕРЕТВОРЕННЯ.....	38
2.1 Підходи до спектрального аналізу сигналів методом Фур'є	38
2.2 Розклад функції в ряд Фур'є.....	38
2.3 Інтегральне перетворення Фур'є.....	45
2.4 Дискретне перетворення Фур'є	55
2.5 Швидке перетворення Фур'є.....	61
2.6 Пряме та обернене z -перетворення, його властивості.....	69
2.7 Зв'язок ДПФ з z -перетворенням.....	72
2.8 Обернене z -перетворення.....	75
2.9 Загальні відомості про вейвлет-перетворення.....	75
Питання для самоперевірки.....	84
РОЗДІЛ 3. КОРЕЛЯЦІЙНИЙ АНАЛІЗ СИГНАЛІВ ТА ЗОБРАЖЕНЬ.....	85
3.1 Основи кореляційного аналізу сигналів та зображень.....	85

3.2 Автокореляційна функція.....	85
3.3 Функція взаємної кореляції.....	88
3.4 Зв'язок між кореляційними функціями і спектрами сигналів.....	90
Питання для самоперевірки.....	92
РОЗДІЛ 4. АНАЛІЗ ВИПАДКОВИХ СИГНАЛІВ.....	93
4.1 Характеристики випадкових сигналів.....	93
4.2 Опис випадкових функцій.....	98
4.3 Спектральні характеристики випадкових процесів.....	108
Питання для самоперевірки.....	113
РОЗДІЛ 5. СИСТЕМИ ОБРОБКИ СИГНАЛІВ.....	115
5.1 Аналогові системи обробки сигналів.....	115
5.2 Аналогові фільтри.....	123
5.3 Дискретні системи обробки сигналів.....	126
5.4 Цифрові фільтри. Класифікація цифрових фільтрів.....	128
5.5 Нерекурсивні фільтри.....	132
5.6 Рекурсивні фільтри.....	136
5.7 Порівняння типів ЦФ.....	140
Питання для самоперевірки.....	141
РОЗДІЛ 6. МОДУЛЯЦІЯ ТА ДЕМОДУЛЯЦІЯ СИГНАЛІВ.....	142
6.1 Модуляція сигналів.....	142
6.2 Демодуляція сигналів.....	153
Питання для самоперевірки.....	154
РОЗДІЛ 7. ОСОБЛИВОСТІ ЦИФРОВОЇ ОБРОБКИ	
ЗОБРАЖЕНЬ.....	156
7.1 Основи цифрової обробки зображень.....	156
7.2 Цифрова обробка кольорових зображень.....	163
Питання для самоперевірки.....	180
Перелік використаних джерел.....	180

ВСТУП

Аналіз та обробка сигналів є одним з найважливіших напрямків розвитку сучасних інформаційних технологій. Фактично в будь-якому технічному засобі, де в тій або іншій формі використовується певний сигнал, присутні елементи обробки сигналів. Відповідно фахівці, задіяні в створенні таких засобів, мають бути обізнаними у необхідних межах з об'єктами вивчення навчальної дисципліни «Цифрова обробка сигналів та зображень». Метою дисципліни є підготовка фахівців, які володіють методами побудови та експлуатації засобів цифрової обробки сигналів та зображень, вміють на основі отриманих теоретичних результатів виконувати як моделювання, так і розробку таких засобів. Дана дисципліна необхідна для формування у студентів знань про методи та засоби передачі, обробки, аналізу сигналів і зображень за допомогою цифрових систем. Розглянутий матеріал дасть змогу майбутньому фахівцю створювати та застосовувати сучасні комп'ютерно-інтегровані технології, розробляти цифрові системи, проводити автоматизацію процесів обробки сигналів і зображень. Набуті під час вивчення дисципліни знання, вміння та досвід можуть бути використані студентами в подальшому при вивченні спеціальних дисциплін.

В представленому в даному посібнику лекційному матеріалі містяться найбільш загальні теоретичні відомості. Разом з інформацією, яка надається студентам на комп'ютерних практикумах та в ході виконання розрахунково-графічної роботи, наданий об'єм знань дозволяє їм самостійно засвоїти повний курс дисципліни.

Матеріал посібника містить сім розділів. В першому розділі розглянуто загальні питання подання сигналів в аналоговому та цифровому видах, представлено основні види тестових сигналів. Другий розділ присвячено спектральному аналізу сигналів в аналоговій та цифровій формах. Вивчаються методи розкладу в ряд Фур'є, пряме та обернене інтегральне перетворення Фур'є, z-перетворення цифрових сигналів. Надано базову інформацію про вейвлет-перетворення. В третьому розділі подано основи кореляційного аналізу сигналів. Розглянуто

автокореляційна та функція взаємної кореляції, зв'язок між кореляційними функціями і спектрами сигналів. Характеристики та елементи аналізу випадкових сигналів викладено в четвертому розділі. В п'ятому розділі розглядаються теоретичні основи обробки аналогових та цифрових сигналів в лінійних системах, особливості аналогових та цифрових фільтрів. Шостий розділ присвячено базовим питанням модуляції та демодуляції сигналів. В сьомому розділі надано стиснуту інформацію про можливості та засоби цифрової обробки зображень. Особливу увагу приділено обробці кольорових зображень.

Навчальний посібник відповідає програмі підготовки магістрів за спеціальністю 174 «Автоматизація, комп'ютерно-інтегровані технології та робототехніка». Також він рекомендується фахівцям, що працюють в проектних та конструкторських організаціях, пов'язаних з розробкою комп'ютеризованих приладів та систем.

РОЗДІЛ 1. КЛАСИФІКАЦІЯ ТА ОСНОВИ ЦИФРОВОГО ПРЕДСТАВЛЕННЯ СИГНАЛІВ

1.1. Вступ до цифрової обробки сигналів та зображень.

Класифікація вимірювальних сигналів

Сигнал – це певна функція (тобто залежність однієї величини від іншої), що містить повідомлення про фізичні властивості, стан або поведінку якої-небудь фізичної системи, об'єкта або середовища. Метою *обробки* сигналів є отримання певних інформаційних відомостей, що відображені в цих сигналах та перетворення цих відомостей у форму, зручну для сприйняття й подальшого використання. Зазвичай розглядаються часові та частотні залежності.

Однак, можливі, наприклад, і просторові залежності (широко застосовуються при геометричних вимірюваннях при оптичній обробці сигналів, в радіолокації, в гідролокації). Фізична природа сигналу також може бути різною. Це може бути напруга, струм, освітленість, яскравість, тиск тощо.

При вимірюваннях сигнал визначають як процес, властивості якого встановлюються взаємодією між матеріальним об'єктом та засобом його дослідження [1].

Параметром сигналу є одна із властивостей сигналу, яка є фізичною величиною.

Інформативний параметр сигналу – це параметр сигналу, який функціонально пов'язаний з досліджуваною або вимірюваною величиною чи той, що має задане значення.

Вимірювальний сигнал – сигнал, один чи кілька параметрів якого є інформативними.

В загальному випадку сигнал може бути багатовимірним, тобто містити упорядковану сукупність одновимірних сигналів, деяким чином пов'язаних між собою. Наприклад, вони можуть одночасно залежати від кількох часових і просторових параметрів, які потребують вимірювання.

За характером вимірювань інформативного і часового параметрів вимірювальні сигнали поділяють на *аналогові*, *дискретні* та *цифрові*

(рис. 1.1) [2].

Аналогові сигнали – описуються неперервною або кусково-неперервною функцією $y_a(t)$, причому ця функція та її аргумент t можуть набувати будь-яких значень на заданих інтервалах $y \in (y_{min}, y_{max})$ і $t \in (t_{min}, t_{max})$. У природі всі вимірювані фізичні величини на макрорівні мають аналоговий характер. Назву «аналоговий» дали неперервним сигналам, тому що вони є «аналогічними» фізичним процесам, які відбуваються в дійсності.

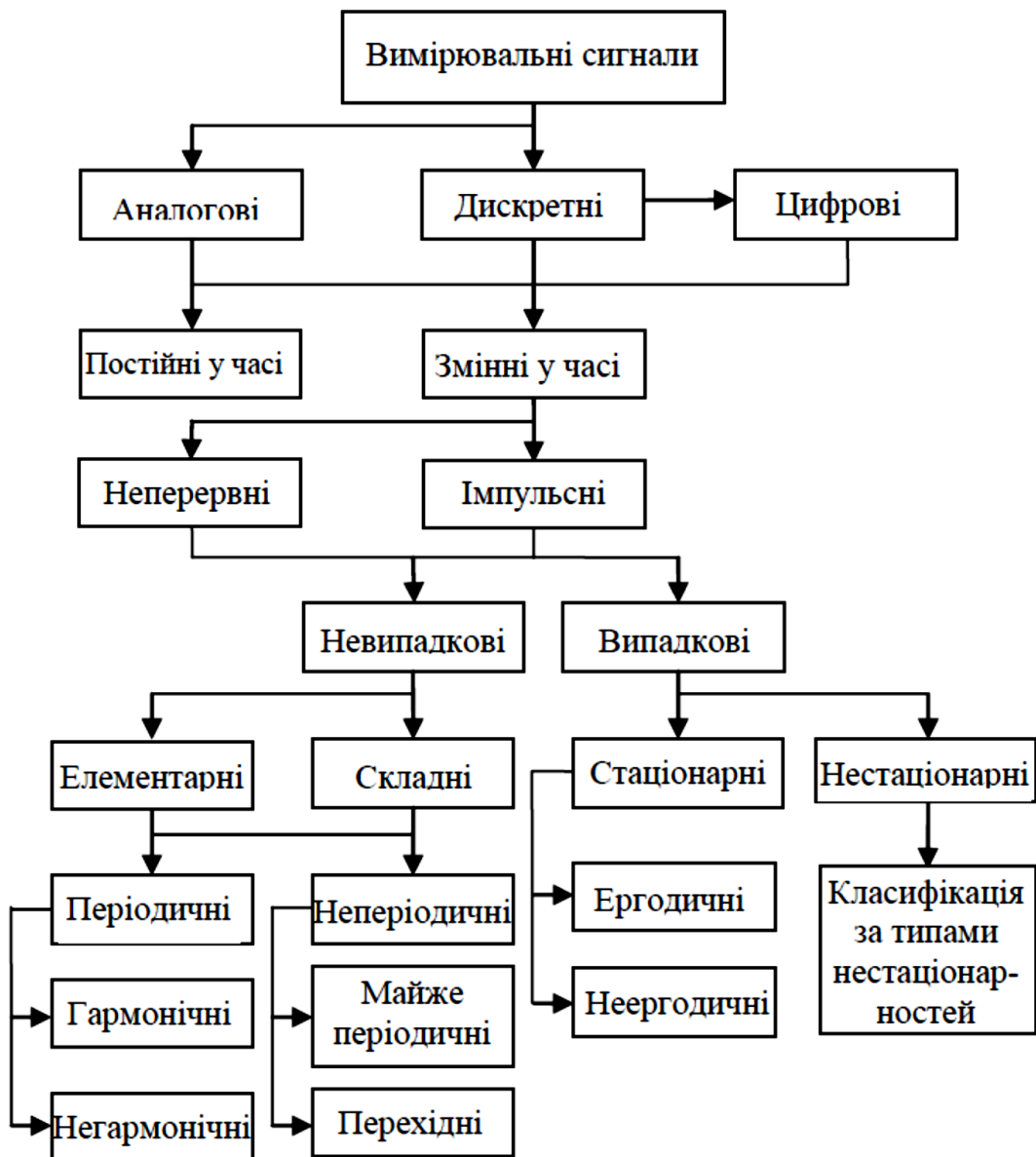


Рис. 1.1. Класифікація вимірювальних сигналів

Сигнали можуть бути об'єктами досліджень і аналізу тільки в тому випадку, якщо зазначено спосіб їхнього математичного опису, тобто існує

математична модель сигналу. Математичний опис дозволяє абстрагуватися від фізичної природи сигналу, проводити класифікацію сигналів, виконувати їхнє порівняння, установлювати ступінь тотожності, моделювати системи обробки сигналів. Як правило, опис сигналу задається функціональною залежністю певного інформаційного параметра сигналу від незалежної змінної (аргументу) – $s(x)$, $y(t)$ тощо.

Теорія аналізу й обробки фізичних даних базується на математичних моделях відповідних фізичних полів і фізичних процесів, на основі яких створюються математичні моделі сигналів. Математичні моделі сигналів дають можливість узагальнено, абстрагуючись від фізичної природи, робити висновок про властивості сигналів, передбачувати зміни сигналів в умовах, що змінюються, замінити фізичне моделювання досліджуваних процесів математичним.

За допомогою математичних моделей є можливість описувати властивості сигналів, які є визначальними в досліджуваних процесах, та ігнорувати велику кількість другорядних ознак. Знання математичних моделей сигналів дає можливість класифікувати їх за різними ознаками, що характерні для того або іншого типу моделей.

Так, сигнали розділяють на не випадкові й випадкові, залежно від можливості точного передбачення їхніх значень у будь-які моменти часу.

Сигнал є не випадковим і називається *детермінованим*, якщо математична модель дозволяє здійснювати таке передбачення. Детермінований сигнал задається, як правило, математичною функцією або обчислювальним алгоритмом, а математична модель сигналу може бути представлена у вигляді

$$s = F(t, z, \omega, \dots; A, B, C, \dots),$$

де s - інформативний параметр сигналу; t, z, ω , незалежні аргументи (час, просторова координата, частота тощо); A, B, C, \dots , - параметри сигналів.

Модель повинна бути, за можливістю, найпростішою, містити мінімальну кількість незалежних аргументів.

Математичний опис не може бути всеосяжним та ідеально точним і, по суті, завжди відображає не реальні об'єкти, а їх спрощені моделі. Моделі

можуть задаватися таблицями, графіками, функціональними залежностями, рівняннями станів і переходів з одного стану в інший тощо.

Формалізований опис може вважатися математичною моделлю оригіналу, якщо він дозволяє з певною точністю прогнозувати стан і поведінку досліджуваних об'єктів шляхом формальних процедур над їхнім описом.

Невід'ємною частиною будь-якої математичної моделі сигналу є область визначення сигналу, що встановлюється інтервалом задання незалежної змінної. Крім задання області визначення сигналу можуть бути також задані види чисельних значень змінних (цілі, раціональні, дійсні, комплексні).

При аналізі фізичних даних використовують два основних підходи до створення математичних моделей сигналів.

Перший підхід оперує з *детермінованими* сигналами, значення яких у будь-який момент часу або в довільній точці простору (а також залежно від будь-яких інших аргументів) є апріорно відомими або можуть бути досить точно визначені. Для опису не випадкових сигналів використовуються також *квазидетерміновані* моделі, у яких значення одного або декількох параметрів апріорно є невідомими та вважаються випадковими величинами з малим випадковим компонентом, впливом якого можна знехтувати.

Другий підхід припускає *випадковий* характер сигналів, закон зміни яких у часі (або в просторі) має випадковий характер і які набувають конкретних значень з деякою імовірністю. Модель такого сигналу є описом статистичних характеристик випадкового процесу шляхом задання законів розподілу ймовірностей, кореляційної функції, спектральної щільностей енергії тощо.

Строго кажучи, детермінованих процесів і детермінованих сигналів, що відповідають їм, у природі не існує. Навіть сигнали, добре відомі на вході в певну ланку інформаційної системи при їх реєстрації завжди є ускладненими випадковими завадами, впливом дестабілізуючих факторів і апріорно невідомими параметрами процесу перетворення інформації.

З іншого боку, модель випадкового поля часто апроксимується

методом суперпозиції (накладення) сигналів відомої форми. Детерміновані моделі можуть використатись також для вивчення чисто випадкових процесів, якщо рівень корисного сигналу в цьому процесі значно є вищим рівня статистичних флуктуацій.

Детерміновані та квазидетерміновані сигнали поділяють на елементарні (прості), що описуються простими математичними формулами, і складні. До елементарних належать постійні та гармонічні сигнали, а також сигнали, що описуються одиничною і δ -функцією. До складних сигналів належать імпульсні і модульовані сигнали.

На вибір математичної моделі в тому або іншому випадку значно впливає також складність математичного апарату обробки сигналів і сформовані традиції інтерпретації результатів спостережень. Не виключається і зміна моделі, як правило, з ймовірнісної в детерміновану, у процесі нагромадження інформації про досліджуване явище або об'єкт.

Дискретні сигнали змінюються стрибкоподібно у часі або за рівнем. У першому випадку вони можуть набувати у дискретні моменти часу nT , де число $T = \text{const}$ – інтервал (період) дискретизації; $n = 0, 1, 2, \dots$ – ціле число, будь-яких значень $y_a \in (y_{\min}, y_{\max})$, які називають *вибірками* або *відліками* (рис. 1.2). У другому випадку значення сигналу $y_a(t)$ існують у будь-який момент $t \in (t_{\min}, t_{\max})$, але вони можуть набувати обмежений ряд станів $h_i = iq$, кратних діапазону q , $i = 0, 1, 2, \dots$



Рис. 1.2. Дискретизація та квантування аналогового сигналу

Під *дискретизацією* розуміють перетворення неперервного сигналу в дискретний, представлений сукупністю відліків, за якими неперервний сигнал може бути відновлений із заданою точністю. Вважатимемо, що відліки сигналу є відліками часу. Дискретизація за часом - це процедура зняття миттєвих значень сигналу $x(t)$ через певні проміжки часу T (рис.1.2).

Миттєві значення $x[nT]$ називаються вибірками, або відліками, час T - періодом (інтервалом, кроком) дискретизації, n вказує порядковий номер відліку. Очевидно, що чим частіше брати відліки (тобто чим меншим є *період дискретизації* T) тим точніше послідовність відліків $x[nT]$ відобразить вихідний сигнал $x(t)$.

Період дискретизації T визначає *частоту дискретизації* f_d , яка дорівнює $f_d = 1/T$. Тобто чим менше T тим вище f_d . З іншого боку, чим вища частота дискретизації, тим важче обчислювачу виконувати велику кількість операцій над відліками в темпі їхнього надходження на переробку і тим складнішою повинна бути його будова.

Аналогом дискретизації для цифрового сигналу є операція проріджування. Суть її полягає в тому, що за певними правилами можна без втрати інформації «відкидати» деякі члени послідовності, тобто проріджувати цифровий сигнал. Операція проріджування називається ще децимацією. Її виконує пристрій, який називається дециматором. Зворотна їй операція відновлення сигналу називається інтерполяцією, її реалізує інтерполятор.

Під *квантуванням* розуміють перетворення деякої величини з неперервною шкалою значень на величину з дискретною шкалою значень. Воно полягає в заміні будь-якого значення відліку одним із кінцевої множини дозволених значень або рівнів квантування.

Квантування виконують з метою формування послідовності чисел $x[nT]=x[n]$, поданих у двійковому коді. Для цього весь діапазон змін величини відліків ділиться на деяку кількість дискретних рівнів і кожному відліку за певними правилами присвоюється значення одного з двох найближчих рівнів квантування, між якими розміщений цей відлік.

Різниця між сусідніми квантованими значеннями визначає *крок квантування*, а різницю між фактичним і квантованим значенням називають *помилкою квантування* ϵ . Максимальне значення цієї помилки дорівнює половині кроку квантування. На підставі потрібного значення помилки квантування вибирають кількість *рівнів квантування*. Що більше рівнів квантування, то менша помилка.

Оскільки цілі числа (рівні квантування) розташовані через однакові інтервали, то говорять про рівномірне квантування. Відомо, що середня помилка рівномірного квантувача випадкового сигналу дорівнює $\Delta^2/12$, де Δ - крок квантувача. Отже, у випадку операції заокруглення середня помилка дорівнює $1/12$. За рівномірного квантування випадкового сигналу співвідношення сигнал/шум на виході квантувача приблизно дорівнює $6N$ дБ, де N - кількість бітів, які використовуються для кодування одного відліку. Наприклад, у 12-розрядного АЦП відношення сигнал/шум може дорівнювати 72 дБ.

Кількість рівнів квантування визначається розрядністю аналого-цифрового перетворювача (АЦП). Так, АЦП на три розряди може мати лише вісім рівнів квантування ($k=2^3$), а мінімальне і максимальне значення відліків дорівнюють відповідно $0 \leftrightarrow 000$ і $7 \leftrightarrow 111$. Чим більшою є розрядність АЦП, тим точнішим є поданий відлік, однак складнішим і дорожчим є АЦП, який необхідний для розв'язання поставленої задачі.

Похибка квантування може призводити до певних проблем. Так, похибка квантування, яка є схожою на сигнал, наприклад, музичний, дає відчуття «брудного» звуку під час прослуховування. Для надання похибці більш випадкового характеру до вихідного аналогового сигналу можна додати деяку кількість високочастотного шуму.

Цей метод називається *дизеризацією*, а відповідний квантувач - дизеризованим. Енергія похибки в разі використання цього методу дещо зростає, але похибка стає не корельованою із сигналом, що у деяких випадках покращує якість звучання.

Найефективніший, але трудомісткий метод декореляції помилки квантування називається *дизеризацією з відніманням*. Його суть така: спочатку генерують і запам'ятовують послідовність псевдовипадкових

чисел; потім її пропускають через АЦП і складають з аналоговим сигналом. На виході АЦП запам'ятовані випадкові числа віднімаються від оцифрованого сигналу.

Зменшити похибку квантування можна, збільшивши кількість рівнів квантування. Однак, це не завжди можливо. Разом з тим, у техніці широко застосовують *компандери*. Це пристрої, які стискають динамічний діапазон аналогового сигналу перед квантувачем і розширюють його після квантувача. Це дозволяє зменшити похибки квантування.

Останнім часом під квантуванням часто розуміють процес перетворення сигналів із дискретною шкалою значень також на цифрові сигнали, діапазон значень яких є меншим. Особливо важливе значення таке перетворення має для стиснення сигналів, тобто зменшення розмірності їх цифрового опису.

Під час цього аналоговий сигнал спочатку перетворюється на цифровий за допомогою квантування достатньо великою кількістю рівнів, яке виконує АЦП, а потім цей сигнал «стискається» в системі цифрової обробки сигналу (ЦОС). Стиснення досягається за рахунок двох речей.

По-перше, цифрові значення відліків реальних сигналів виявляються корельованими. Так, коефіцієнт кореляції сусідніх відліків мови приблизно дорівнює 0,9. Ця кореляція може застосовуватись, наприклад, у диференційних системах кодування або в системах стиснення з прогнозуванням.

По-друге, при стисненні використовуються властивості людського слуху (зору), який подібний на фільтр, що пропускає певні частоти (та типи спотворень) і затримує інші. Завдяки цим властивостям часто спостерігається ефект розбіжності між суб'єктивно спостережуваною якістю сигналу та об'єктивно виміряною мірою спотворення.

Вибір міри спотворення є одним із найважливіших при проектуванні квантувача. Найчастіше використовують середньоквадратичну похибку, яка виражається в зручній математичній формі і багато в чому відображає властивості чуттєвого сприйняття.

Інколи квантування безпосереднього сигналу не настільки ефективне, як квантування його перетвореної версії. Причиною цього є те, що за

належного вибору перетворення, його коефіцієнти можуть виявитися декорельованими. Зазвичай використовують ортогональні перетворення, наприклад, Фур'є, дискретне косинусне, вейвлет-перетворення.

Найчастіше кодування з перетворенням застосовують для стиснення зображень. Так, в основу стандарту JPEG покладено дискретне косинусне перетворення, а в основу стандарту JPEG2000 - вейвлет-перетворення зображення.

На завершальній стадії багатьох алгоритмів стиснення використовується ентропійний кодер. Таким кодером найчастіше слугує поєднання кодера довжин серій з кодером Хафмана або арифметичним кодером.

Під час побудови алгоритму стиснення існують дві можливості: аналізувати і квантувати кожний відлік сигналу окремо або разом з іншими відліками, об'єднуючи їх у вектор. Відповідно розрізняють скалярне та векторне квантування.

Скалярний квантувач характеризується двома параметрами: кількістю рівнів квантування та кроком квантування. Найпростішим видом скалярного квантувача є рівномірний квантувач.

У результаті квантування, яке здійснюється аналого-цифровим перетворювачем, у канал зв'язку передається послідовність двійкових чисел. Кількість таких чисел визначається частотою дискретизації f_d , а швидкість C їх передавання в каналі залежить ще і від розрядності b представлення чисел $C = bf_d$, і під час передавання по каналу, наприклад, мовного сигналу за стандартних $b = 12$, $f_d = 8000$ Гц отримуємо $C = 96000$ біт/с, що перешкоджає використанню стандартних телефонних каналів.

У таких випадках використовують *алгоритми* стиснення сигналу.

Цифрові сигнали – це квантовані за рівнем і дискретні за часом сигнали $Y(nT)$, що описуються квантованими решітчастими функціями (квантованими послідовностями), які набувають у дискретні моменти часу nT лише скінченний ряд дискретних значень – рівнів квантування h_1, h_2, \dots, h_l (рис. 1.3). Характерним для таких сигналів є представлення їх відліків за рівнем у формі чисел, зазвичай двійкової розрядності.

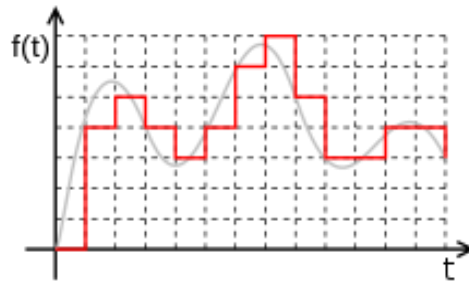


Рис. 1.3. Цифровий сигнал

На рис. 1.4 показано описані типи сигналів [2].

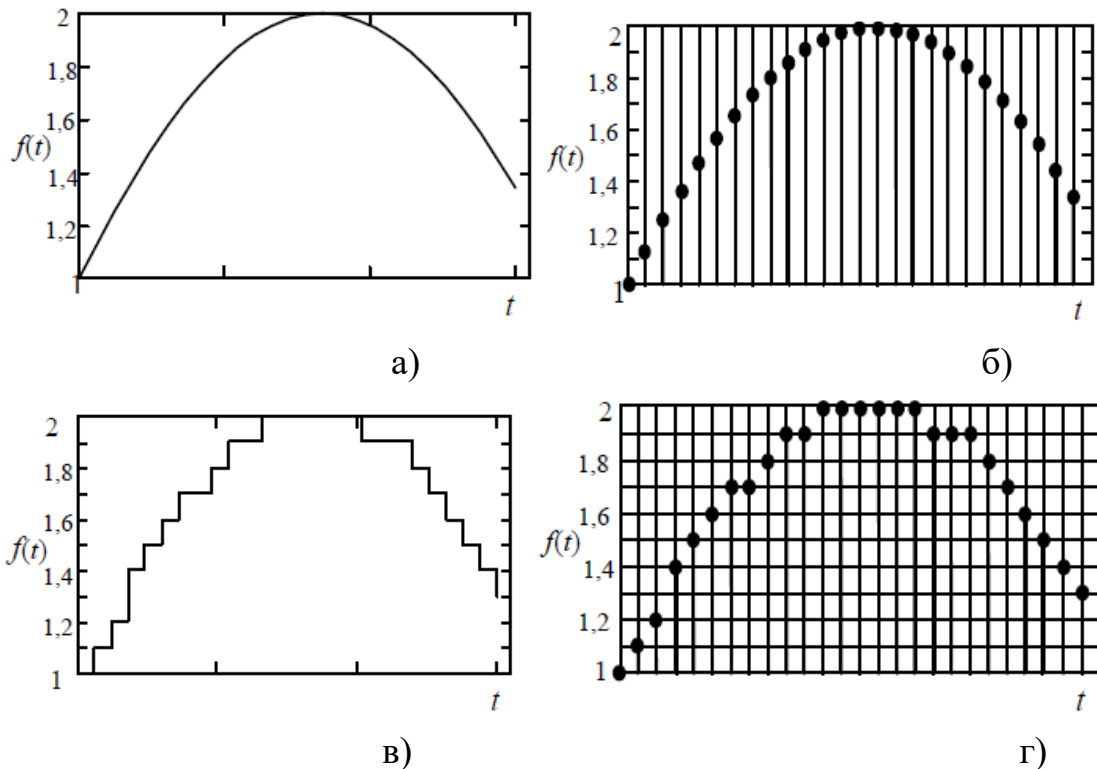


Рис. 1.4. Вимірювальні сигнали: а - аналоговий; б - дискретний за часом; в - квантований за рівнем; г - цифровий

За характером зміни у часі сигнали поділяють на *постійні*, з незмінними у часі значеннями, і *змінні*, значення яких змінюються з часом. Постійні сигнали є найпростішим видом вимірювальних сигналів. Змінні сигнали можуть бути *неперервними* за часом та *імпульсними*.

Неперервним називають сигнал, який існує на нескінченній множині точок часової осі і триває нескінченно довго ($-\infty < t < \infty$). Параметри такого сигналу змінюються неперервно.

Імпульсні сигнали – це сигнали кінцевої енергії, істотно відмінні від

нуля впродовж обмеженого інтервалу часу, який приблизно дорівнює часу завершення перехідного процесу в системі, для впливу на яку цей сигнал призначений. Іноді в окремий клас виділяють сигнали кінцевої тривалості, відмінні від нуля тільки на обмеженому інтервалі аргументів. Такі сигнали звичайно називають *фінітними*.

Сигнали можуть бути періодичними і неперіодичними.

Періодичними називають сигнали, миттєві значення яких повторюються через постійний інтервал часу $u(t) = u(t + nT)$,

де n – довільне ціле число; T - період сигналу (це параметр, який дорівнює найменшому такому інтервалу часу); частота періодичного сигналу – є величиною, оберненою періоду $f = \frac{1}{T}$.

Прикладом періодичного сигналу є положення проекції точки P , яка рівномірно обертається по колу з кутовою частотою ω , на вісь ординат (рис. 1.5), або прямокутний, пилкоподібний, трикутний сигнали (рис. 1.6) [3].

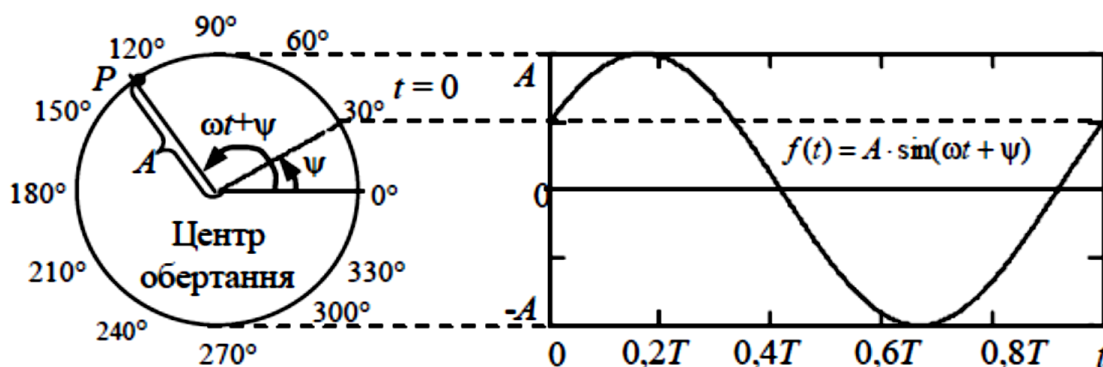


Рис. 1.5. Положення проекції точки, яка обертається, на вертикальну вісь

Періодичні сигнали бувають гармонічними (такими, що мають тільки одну гармоніку) і негармонічними (полігармонічними), спектр яких складається з множини гармонічних складових. До гармонічних сигналів належать сигнали, що описуються функцією синуса або косинуса

$$s(t) = A \cdot \sin(2\pi f_0 t + \psi) = A \cdot \sin(\omega_0 t + \psi),$$

$$s(t) = A \cdot \cos(\omega_0 t + \psi),$$

де A - амплітуда сигналу; f_0 - лінійна частота в Гц; $\omega_0 = 2\pi f_0$ - кругова частота в радіанах; ψ - початковий фазовий кут в радіанах.

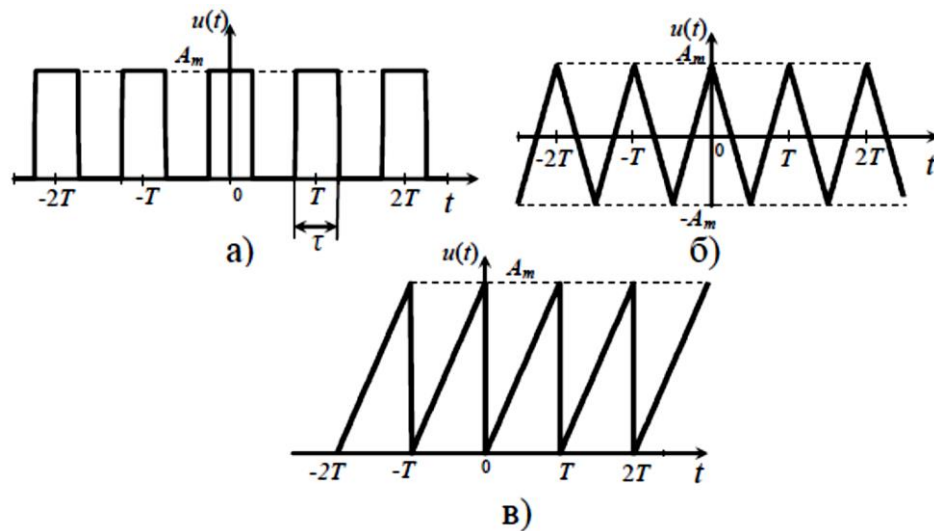


Рис. 1.6. Приклади періодичних сигналів:
 а - прямокутний; б - трикутний; в - пилкоподібний

Період одного коливання сигналу становить $T=1/f_0=2\pi/\omega_0$.

Полігармонічні сигнали складають найбільш поширену групу періодичних сигналів і описуються сумою певної постійної складової ($f_0=0$) і гармонічних коливань з довільними значеннями амплітуд A_n і фаз φ_n

$$s(t) = \sum_{n=0}^N A_n \sin(2\pi f_n t + \varphi_n),$$

або безпосередньо функцією $s(t) = y(t \pm kT_p)$, $k = 1, 2, 3, \dots$, де T_p - період одного повного коливання сигналу $y(t)$, заданого на одному періоді. Величину $f_p=1/T_p$ називають фундаментальною частотою коливань.

До *неперіодичних* сигналів відносять *майже періодичні* і *перехідні*. Майже періодичними називають сигнали, значення яких приблизно повторюються при додаванні до часового аргументу належним чином вибраного числа – майже періоду. Прикладом такого сигналу є сума періодичних функцій з некрatними періодами, наприклад, $u(t)=\cos(\omega_0 t)+\cos(3\omega_0 t)$. Майже періодичні функції, як і періодичні, також мають дискретний спектр, але тільки амплітудний з некрatними частотами.

Перехідні сигнали описують перехідні процеси у фізичних системах. Перехідні неперіодичні сигнали мають неперервні спектри. Наприклад для прямокутного імпульсу виду $\sin(x)/x$.

Випадковий сигнал – це змінна у часі величина, миттєве значення якої є випадковою величиною. Різновидом випадкових сигналів є *завади*. Під

завадою (*перешкодою*) розуміють сигнал однорідний з вимірювальним, що діє одночасно з ним. Наявність завад приводить до появи додаткової похибки (невизначеності) вимірювання [2].

Між сигналами і завадами немає принципової різниці. Можлива ситуація, коли корисний сигнал для однієї вимірювальної системи є завадовим для іншої (наприклад, в системах стільникового зв'язку при повторному використанні частот в ретрансляційних пунктах; в радіолокаційних системах, які працюють на однакових частотах, випромінювані однією радіолокаційною станцією (РЛС) сигнали утворюють несинхронну імпульсну заваду для інших РЛС). Тобто, наявність у вимірювальній системі каналу передавання робить принципово можливим утворення для неї завад сигналами подібними до тих, що передаються цими каналами.

За місцем виникнення завади поділяють на зовнішні та внутрішні. Причиною виникнення зовнішніх завад є природні процеси і робота різних технічних пристроїв. Останні створюють так звані індустриальні завади.

Природними джерелами зовнішніх завад (шумів) є: випромінювання позаземних джерел або шум космічного простору; шум, зумовлений флуктуаційним характером поглинання радіохвиль в атмосфері Землі; шум, зумовлений тепловим випромінюванням Землі; завади від грозових розрядів в атмосфері тощо.

Основними джерелами індустриальних завад є: працюючі електричні апарати (високочастотна апаратура для дугового зварювання, газорозрядні пристрої, динамо-машини, генератори електричного струму, індукційна й комутаційна апаратура); системи запалювання двигунів внутрішнього згоряння різного призначення; лінії електропередач.

Внутрішні завади зумовлені процесами, що виникають під час роботи самої вимірювальної системи.

За основними властивостями завади можна поділити на три види: флуктуаційні, зосереджені та імпульсні. *Флуктуаційні* завади відображають хаотичну, випадкову зміну з часом сигналу, однорідного з вимірюваним, у будь-якому місці вимірювального пристрою. Такі завади часто називають *шумом*. Вплив флуктуаційної завади може бути

зменшений для періодичних ВС при накопиченні сигналу у кількох періодах (завдяки випадковому характеру завади з нульовим математичним очікуванням при складанні достатньо великої кількості періодів сигналу середнє значення завади наближається до нуля).

Зосередженими називають завади, основна частина потужності яких зосереджена на окремих ділянках діапазону частот роботи ЗВ. Завади ЗВ від промислової силової мережі частотою 50 Гц є зосередженими. Ефективність їх заглушення великою мірою визначається наявністю апріорних даних про частотний спектр.

Імпульсними завадами називають регулярну або хаотичну послідовність імпульсних сигналів, які є однорідними з вимірвальним сигналом. Джерелами таких завад є цифрові комутаційні елементи ЗВТ або працюючого поряд з ними пристрою.

Характерний приклад імпульсних завад – завади від пристроїв запалювання двигунів внутрішнього згорання (групи імпульсів з випадковими амплітудами і часовими інтервалами тривалістю в межах від кількох мікросекунд до кількох мілісекунд).

З енергетичних позицій сигнали розділяють на два класи: з обмеженою (кінцевою) енергією та з нескінченною енергією.

Для сигналів з обмеженою енергією (сигналів, що інтегруються з квадратом) повинне виконуватися співвідношення скінченності енергії сигналу, тобто її обмеженості. Як правило, до цього класу сигналів відносять аперіодичні й імпульсні сигнали, що не мають розривів 2-го роду при обмеженій кількості розривів 1-го роду, і особливих точок, що йдуть у нескінченність.

Будь-які періодичні, полігармонічні та майже періодичні сигнали, а також сигнали з розривами й особливими точками 2-го роду, є сигналами з нескінченною енергією.

1.2. Математичне представлення сигналів. Тестові сигнали

Найбільш поширеними елементарними тестовими сигналами, окрім гармонічного сигналу, є одинична функція (синоніми: функція Хевісайда, функція включення) і δ -функція (функція Дірака).

Такі сигнали застосовуються для динамічного (тобто – в часі) представлення сигналів і обчислення реакції фізичної системи на відомий вхідний сигнал.

Функція Хевісайда (рис. 1.7) описується рівнянням

$$\sigma(t - t_0) = \begin{cases} 0, & t < t_0, \\ \frac{1}{2}, & t = t_0, \\ 1, & t > t_0. \end{cases}$$

Вона має один параметр – момент часу t_0 .

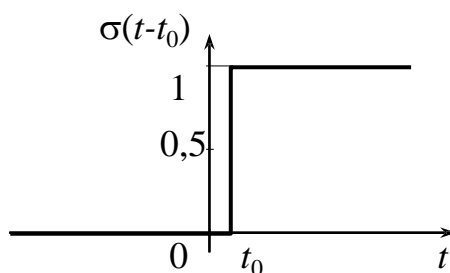


Рис. 1.7. Функція Хевісайда

Функція Хевісайда може бути застосована для наближеного представлення довільного сигналу $s(t)$ ступінчастою функцією вигляду рис. 1.4, в):

$$s(t) \approx s_0 \sigma(t) + (s_1 - s_0) \sigma(t - \Delta t) + (s_2 - s_1) \sigma(t - 2\Delta t) + \dots = s_0 \sigma(t) + \sum_{k=1}^{\infty} (s_k - s_{k-1}) \sigma(t - k\Delta t).$$

Якщо крок дискретизації Δt спрямувати до нуля, то $k\Delta t$ буде наближатись до неперервної змінної t' , а малі прирости сигналу $(s_k - s_{k-1})$ представлятимуть диференціали $ds = \frac{ds}{dt'} dt'$. Додавання при цьому перетвориться в інтегрування в границях від нуля до нескінченності. В результаті отримаємо динамічне (у часі) представлення довільного сигналу у вигляді

$$s(t) = s_0 \sigma(t) + \int_0^{\infty} \frac{ds}{dt'} \sigma(t - t') dt'.$$

Прямокутна функція (рис. 8) описується як

$$\text{rect}(t) = \begin{cases} 1, & \text{при } |t| \leq \frac{1}{2} \\ 0, & \text{при } |t| > \frac{1}{2} \end{cases}.$$

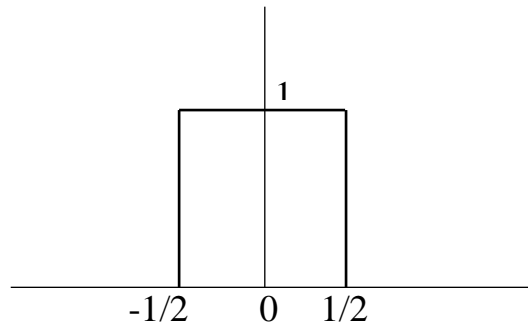


Рис. 1.8. Прямокутна функція

Поодинокий ідеальний прямокутний імпульс тривалістю τ з амплітудою A_m , який з'являється в момент часу t_0 , описується рівнянням

$$s(t) = A_m[\sigma(t - t_0) - \sigma(t - t_0 - \tau)].$$

Тобто він формується як різниця двох одиничних функцій, зсунутих за часом на тривалість імпульсу τ (рис. 1.9).

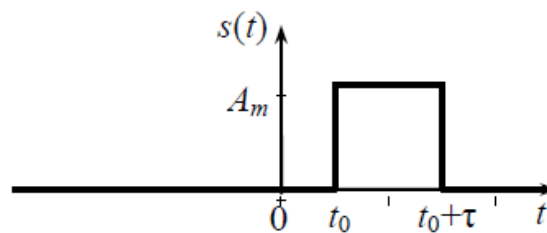


Рис. 1.9. Прямокутний імпульс

δ -функція (функція Дірака) описується рівнянням

$$\delta(t - t_0) = \begin{cases} 0, & \text{при } t \neq t_0, \\ \infty, & \text{при } t = t_0. \end{cases}$$

δ -функція є імпульсом нескінченно короткої тривалості з нескінченно великою амплітудою та площею, що дорівнює одиниці. Приклади зображення δ -функції подані на рис. 1.10.

Фізично δ -функцію можна визначити як граничне наближення

$$\delta(t) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \text{rect}(t).$$

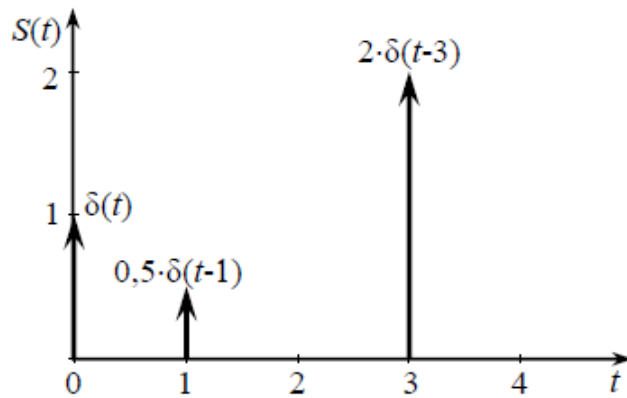


Рис. 1.10. Графік сигналу $s(t) = \delta(t) + 0,5 \delta(t-1) + 2 \delta(t-3)$

Функція Дірака має такі властивості:

- Її площа дорівнює одиниці: $\int_{-\infty}^{\infty} \delta(t - t_0) dt = \int_{t_0 - \varepsilon}^{t_0 + \varepsilon} \delta(t - t_0) dt = 1$,

де ε - будь-яке число, більше нуля.

- Одинична і δ -функції пов'язані між собою співвідношеннями

$$\sigma(t - t_0) = \int_0^t \delta(t - t_0) dt; \quad \delta(t - t_0) = \frac{d\sigma(t - t_0)}{dt}.$$

Тобто δ -функція є похідною від функції Хевісайда, а функція Хевісайда – інтегралом від δ -функції.

- Фільтрувальна властивість, яка описується рівнянням

$$\int_{-\infty}^{\infty} s(t) \delta(t - t_0) dt = s(t_0)$$

Остання властивість використовується для динамічного представлення сигналу

$$s_{\Delta t}(t) = \sum_{n=-\infty}^{\infty} s(n\Delta t) \delta(t - n\Delta t)$$

яке відповідає апроксимації сигналу зімкненою послідовністю прямокутних імпульсів при спрямуванні їх тривалості до нуля, а також дискретизованої у часі функції з кроком дискретизації Δt .

Гармонічний сигнал описується наведеним вище рівнянням синусоїди $f(t) = A \cdot \sin(\omega t + \psi)$ або косинусоїди $f(t) = A \cdot \cos(\omega t + \psi)$.

Параметрами такого сигналу є: амплітуда A , кругова частота $\omega = 2\pi/T$ (період T або частота $f = 1/T$) і початкова фаза ψ .

Треба розрізняти *миттєве значення* сигналу і його *амплітуду*. Миттєве значення – це міра, що визначає на яку величину і в якому напрямку змінна

відхиляється від нуля. Це значення може бути як додатним, так і від'ємним. Модуль змінної показує, наскільки її величина відхиляється від нуля незалежно від напрямку відхилення. Амплітудою називається модуль найбільшого відхилення змінної від нуля. Тобто амплітуда завжди є додатною. Потужність сигналу є пропорційною (а іноді – дорівнює) його амплітуді.

1.3. Узагальнені схеми цифрових систем

Розглянемо перетворення типів сигналів для отримання з аналогового – цифрового і навпаки. Такі перетворення застосовуються в цифровій обробці сигналів (ЦОС) (рис. 1.11) [4].

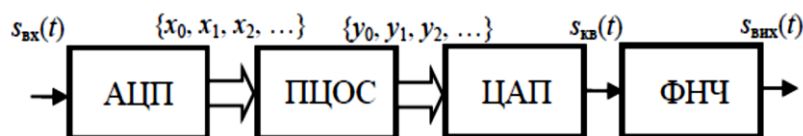


Рис. 1.11. Узагальнена структурна схема системи цифрової обробки сигналів

Часова дискретизація вхідного аналогового сигналу $s_{\text{вх}}(t)$ та його квантування за рівнем виконуються в аналого-цифровому перетворювачі (АЦП). Аналоговий фільтр низьких частот (ФНЧ, low-pass filter) для попередньої обробки сигналу (на рис. 1.11 не показаний) і АЦП разом утворюють пристрій, що називається *кодером*.

На виході АЦП кодера сигнал є послідовністю чисел $\{x_0, x_1, x_2, \dots\}$ певної розрядності, яка поступає в процесор цифрової обробки сигналів (ПЦОС), де виконується необхідна його обробка. В процесорі реалізуються потрібні математичні операції над вхідними відліками; раніше отримані відліки і проміжні результати можуть зберігатись в пам'яті процесора для використання в наступних обчисленнях.

Цифрові відліки вихідного сигналу ПЦОС $\{y_0, y_1, y_2, \dots\}$ за потреби можуть бути використані для отримання аналогового вихідного сигналу за допомогою цифро-аналогового перетворювача ЦАП $s_{\text{кв}}(t)$ і (за необхідності) згладжувального ФНЧ. Пристрій, що об'єднує ЦАП і згладжувальний ФНЧ, називається *декодером*.

Декодер за отриманим цифровим сигналом формує аналоговий сигнал

$s_{\text{вих}}(t)$, тобто виконує обернене порівняно з кодером перетворення. Сигнал $s_{\text{вих}}(t)$ наближається до вигляду аналогового сигналу, що міг би бути отриманий у разі обробки сигналу $s_{\text{вх}}(t)$ аналоговими пристроями.

За допомогою ЦОС можуть бути здійснені усі відомі операції над сигналами. Наприклад, на основі цифрової обробки сигналів в метрології, медицині, радіотехніці вирішуються наступні завдання: генерація, модуляція, формування і корекція сигналів, фільтрація і оцінювання, аналіз спектрів, виявлення і розпізнавання сигналів, кодування і декодування, обробка зображень тощо.

Перевагами ЦОС порівняно з аналоговою обробкою є:

- багатофункціональність обробки, тобто можливість використання одного і того ж процесора для виконання різних операцій обробки шляхом зміни програми;

- мультиплексування, тобто можливість обробки сигналів від декількох датчиків в проміжках між суміжними тактами, що особливо актуально при реалізації багатоканальних інформаційно-вимірювальних систем;

- відсутність принципів обмежень на складність обробки (принципово можливо забезпечити необхідну точність обробки при будь-якій складності алгоритмів);

- зняття обмежень на види обробки, які властиві аналоговій техніці, тобто можливість реалізації будь-яких формально описуваних перетворень.

Реалізація ЦОС в сучасних вимірювальних системах виконується із застосуванням цифрових сигнальних процесорів (ЦСП, digital signal processor, DSP; сигнальний мікропроцесор - СМП; процесор цифрових сигналів - ПЦС; цифровий процесор обробки сигналів - ЦПОС) - спеціалізованих мікропроцесорів, призначених для цифрової обробки сигналів (зазвичай в реальному масштабі часу).

Особливості архітектури сигнальних процесорів порівняно з мікропроцесорами (МП) загального застосування пов'язані з прагненням максимально прискорити виконання типових задач ЦОС: цифрова фільтрація, перетворення Фур'є, кореляційний аналіз тощо. Математично

ці завдання зводяться до поелементного перемножування багатоеlementних векторів дійсних чисел і подальшого підсумовування добутків (наприклад, при цифровій фільтрації вихідний сигнал фільтра з кінцевою імпульсною характеристикою дорівнює сумі добутків коефіцієнтів фільтра на вектор відліків сигналу). Тому ЦСП оптимізовані за швидкодією для виконання саме таких операцій:

1. Операція «множення з накопиченням» (Multiply-Accumulate, MAC) ($Y = Y + A \times B$), де Y , A , B – елементи масивів дійсних чисел, з автоматичним розрахунком адрес елементів масивів, зазвичай реалізована апаратно і виконується за один машинний цикл (рівнозначні 2-5 звичайним командам мікропроцесорів і мікроконтролерів);

2. Апаратна реалізація багаторазового повторення заданого набору команд, тобто циклів заздалегідь призначеної довжини без використання лічильників циклу і команд перевірки обнуління лічильника циклу – ознаки виходу з циклу;

3. Можливість одночасної в одному машинному такті вибірки команди і двох операндів для максимально швидкого виконання команди MAC. Для цього ЦСП має кілька портів звернення до пам'яті (незалежних областей пам'яті, кожна зі своїм комплектом шин адреси і даних);

4. Підтримка векторно-конвеєрної обробки за допомогою генераторів адресних послідовностей.

Оцифровані відліки вхідного сигналу записуються в апаратно організовані кільцеві буфери. Кільцевий буфер схожий на стек, але на відміну від стека він ніколи не переповнюється.

Зміна покажчиків в кільцевому буфері на запис в буфер (стирання старих записів, додавання нових) і читання виконується апаратно. Елементи кільцевого буфера можуть бути аргументами MAC-команди, причому перехід до наступного елементу кільцевого буфера відбувається автоматично за покажчиками при виконанні поточної MAC-команди.

При роботі з даними зовнішньої пам'яті застосовується прямий доступ до пам'яті (ПДП), для чого використовується контролер ПДП або контролер кільцевого буфера з ПДП. Для асинхронного запису і читання в буфер часто використовується двохпортова пам'ять.

Для роботи з синфазними каналами в деяких ЦСП існує два ядра й два набори буферів і контролерів керування кільцевими буферами.

Теорія ЦОС вивчає принципи і методи побудови алгоритмів, програм і технічних засобів цифрової обробки сигналів. Залежно від засобів реалізації (апаратні, програмні або змішані) розглядаються завдання розробки і реалізації алгоритмів ЦОС, методи синтезу фізично можливих операторів для різноманітних прикладних завдань, процедури для досягнення потрібної точності виконання алгоритмів.

Як програмний, так і апаратний способи реалізації ЦОС базуються або на часовому, або на спектральному підході. Розглянемо спочатку принципи часового підходу. Цей підхід потребує, в першу чергу, коректного перетворення аналогового сигналу в дискретний цифровий, тобто представлення його сукупністю відліків у часі, яка є адекватною вхідному сигналу.

1.4. Аналого-цифрове перетворення сигналів

Представлення неперервного сигналу $s(t)$ на інтервалі $[T_1, T_2]$ вектором відліків (координат) $(c_1, c_2, \dots, c_N)^T$ («Т» – операція транспонування) може бути записане у вигляді

$$(c_1, c_2, \dots, c_N)^T = D[s(t)],$$

де D – оператор дискретного представлення сигналу, що реалізовується пристроєм, званим *дискретизатором*.

Аналогічно можна записати операцію відновлення функції $u(t)$, так званою *відтворюючою* функцією, за вектором відліків $(c_1, c_2, \dots, c_N)^T$

$$u(t) = D^{-1}[(c_1, c_2, \dots, c_N)],$$

де D^{-1} – оператор відновлення, обернений до оператора D , який реалізовується пристроєм відновлення сигналу.

Очевидно, відтворююча функція $u(t)$ відображає початковий сигнал $s(t)$ з деякою поточною похибкою наближення (відновлення): $\varepsilon(t) = s(t) - u(t)$.

Завдання дискретизації в математичному плані зводиться до спільного вибору пари операторів D і D^{-1} , що забезпечують задану

точність відновлення сигналу.

У загальному випадку лінійної відтворюючої функції вона може бути представлена апроксимуючим поліномом

$$u(t) = \sum_{i=1}^N c_i \varphi_i(t), \quad (1.1)$$

де $\{\varphi_i(t)\}_{i=1}^N$ – система базисних функцій.

Частіше за все обирають ортогональні базисні функції, для яких скалярний добуток

$$(\varphi_i, \varphi_j) = \int_{-\infty}^{\infty} \varphi_i(t) \varphi_j(t) dt = \begin{cases} 0, & \text{якщо } i \neq j; \\ c, & \text{якщо } i = j, \end{cases}$$

де $c = \text{const}$.

Якщо $c = 1$, базисні функції до того ж є ортонормованими. Такі базисні функції забезпечують збіжність ряду $\sum_{i=1}^N c_i \varphi_i(t)$ до функції $s(t)$ при $N \rightarrow \infty$. (Вектори \mathbf{x} та \mathbf{y} називаються ортогональними, якщо їх скалярний добуток $\mathbf{x}\mathbf{y}=0$. Набір векторів називається ортогональним, якщо будь-які два різних вектори з цього набору ортогональні. Ортогональний набір векторів називається ортонормованим, якщо довжини всіх векторів цього набору дорівнюють 1).

У разі відновлення аналогового сигналу за дискретними відліками з'являється можливість обмежити їх кількість відповідно до заданої похибки відновлення.

Основні завдання, що вирішуються при дискретизації сигналу:

- вибір кроку дискретизації при заданій похибці дискретизації;
- оцінка істотності відліків, що забезпечують представлення сигналу $s(t)$ із заданою точністю, і усунення можливої надмірності.

Алгоритм рішення задачі дискретизації:

- по наявних відліках, використовуючи базисні ортогональні функції і методи інтерполяції або екстраполяції, будують з тією або іншою мірою наближення початковий сигнал, тобто отримують відтворюючу функцію $u(t)$;
- оцінюють ефективність дискретизації при вибраному методі апроксимації, визначаючи похибку дискретизації (відновлення) із заданим кроком дискретизації або крок дискретизації при заданій похибці.

Вибір системи базисних функцій у складі апроксимуючого полінома (1.1) визначається вимогою забезпечення простоти технічної реалізації апаратних (програмних) засобів дискретизації і відновлення сигналу. При неортогональних представленнях сигналу часто застосовують степеневі алгебраїчні поліноми виду

$$u(t) = \sum_{i=0}^N a_i t^i,$$

де a_i - дійсні коефіцієнти.

Як ортогональні базисні функції використовують функції відліків, поліноми Лежандра, функції Уолша тощо. Якщо базисні функції вибрані так, що значення апроксимуючого полінома співпадають зі значеннями вибірки початкового сигналу в моменти їх відліку, то такий поліном називають інтерполяційним. З точки зору скорочення числа відліків інтерполяційні методи відновлення сигналу є більш прийнятними, проте для їх реалізації потрібна затримка сигналу на інтервал інтерполяції, що у ряді випадків є неприпустимим.

Наприклад, у системах управління, що працюють в реальному часі, використовують екстраполяційні методи, які не вимагають затримки сигналу при проведенні операцій знаходження значень вибірок і відновлення сигналу (*екстраполяція* – (від лат. *extra* – ззовні, по за, крім і лат. *polire* – пригладжую, виправляю, змінюю) – особливий тип апроксимації, коли функція апроксимується за межами інтервалу, а не між заданими значеннями. Інакше, екстраполяція – наближене визначення значень функції в точках, що не належать відрізу, по значенням в точках у його середині).

При обраному способі побудови відтворюючої функції повинна забезпечуватись задана точність відновлення сигналу. Залежно від цільового призначення отримуваної інформації використовують різні критерії точності наближення $u(t)$ до $s(t)$, що визначають чисельні значення похибки відновлення і, отже, дискретизації.

Для оцінки точності відновлення (похибки дискретизації) можуть застосовуватись наступні критерії:

1. *Критерій найбільшого відхилення* - встановлює максимальне

значення допустимої похибки відновлення ε_m

$$\varepsilon_m = \max_{t \in \Delta_i} |\varepsilon(t)| = \max_{t \in \Delta_i} |s(t) - u(t)|,$$

де Δ_i - ділянка апроксимації.

2. *Критерій середньоквадратичного наближення* – задає значення допустимої середньоквадратичної похибки відновлення σ_∂

$$\sigma_\partial \geq \sigma = \sqrt{\frac{1}{\Delta_i} \int_{\Delta_i} \varepsilon^2(t) dt},$$

де σ – середньоквадратична похибка наближення. Зазвичай цей критерій використовується для функцій, інтегрованих в квадраті.

3. *Імовірнісний критерій* – визначає допустиму ймовірність P_0 того, що поточна похибка наближення $\varepsilon(t)$ не перевищить деякого певного значення ε_0

$$P[\varepsilon(t) \leq \varepsilon_0] = P_0.$$

Важливою ознакою процесу дискретизації є регулярність відліків. Якщо крок дискретизації постійний в усьому діапазоні перетворення, дискретизація вважається рівномірною. Методи рівномірної дискретизації застосовуються найширше.

Проте невідповідність кроку дискретизації характеру поведінки конкретної реалізації досліджуваного процесу на окремих ділянках часто призводить до значної надмірності відліків.

Якщо крок дискретизації змінюється, наприклад, залежно від швидкості зміни сигналу або за заданою програмою, дискретизацію називають нерівномірною.

Розрізняють *адаптивні* методи нерівномірної дискретизації, при яких крок Δt_i змінюється залежно від поточної зміни параметрів сигналу, і *програмовані* методи, в яких зміна кроку дискретизації виробляється оператором на основі аналізу інформації, що поступає, або відповідно до заздалегідь встановленої програми.

Адаптивні методи нерівномірної дискретизації забезпечують ефективне усунення надмірності у відліках сигналів. Інтервал дискретизації в цьому випадку тісно пов'язаний з поточними значеннями параметрів реалізації сигналу.

1.5. Цифро-аналогове перетворення сигналів

Дискретизація передбачає можливість подальшого відновлення аналогового сигналу із дискретного. Але тут виникає невизначеність.

Нехай в результаті рівномірної дискретизації синусоїдального сигналу отримані такі відліки (рис. 1.12) [5]

$$x(0) = 0; x(1) = 0,866; x(2) = 0,866; x(3) = 0; x(4) = -0,866;$$
$$x(5) = -0,866; x(6) = 0$$

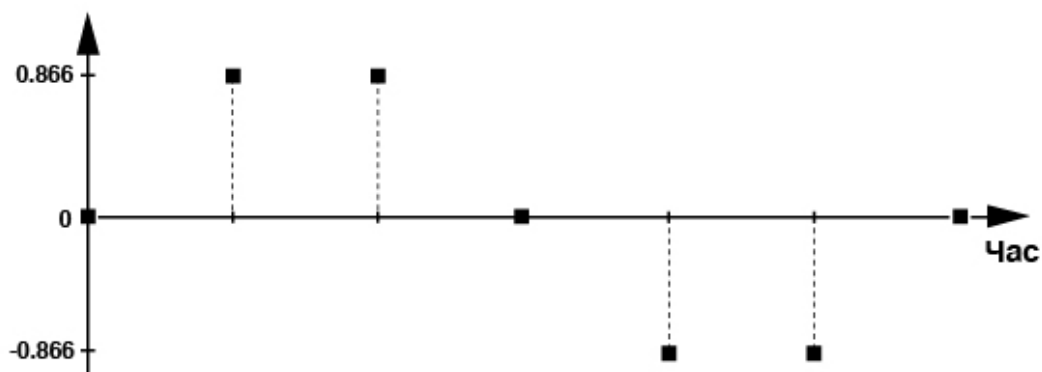


Рис. 1.12. Дискретні відліки

За цими відліками потрібно відновити аналоговий сигнал. Він має проходити через ці точки. Як видно на рис. 1.13, через вказані точки можна провести дві синусоїди.

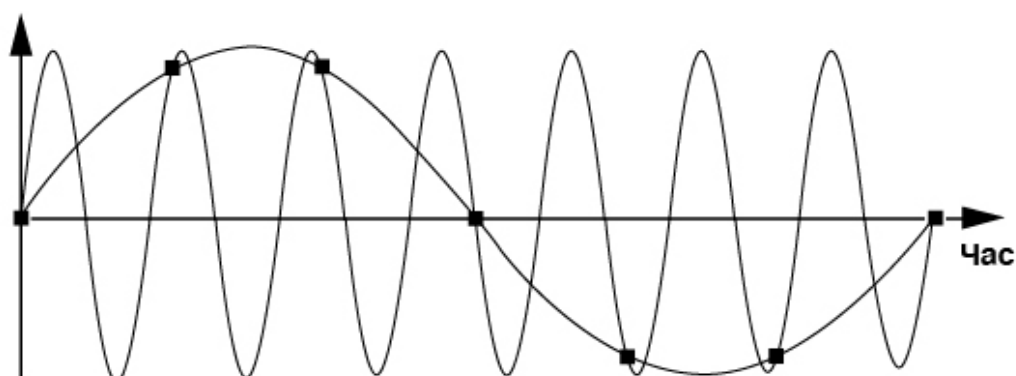


Рис. 1.13. Відновлення синусоїдальних сигналів після дискретизації

Встановимо, чому так сталося. Розглянемо безперервний синусоїдальний сигнал частотою f_0

$$x(t) = \sin(2\pi f_0 t).$$

Виконаємо дискретизацію цього сигналу частотою f_D , тобто через однакові проміжки часу $t_D = 1/f_D$ секунд. В результаті отримуємо n відліків в моменти часу $0t_D, 1t_D, 2t_D, \dots, nt_D$:

$$\begin{aligned} x(0) &= \sin(2\pi f_0 0t_D); \\ x(1) &= \sin(2\pi f_0 1t_D); \\ x(2) &= \sin(2\pi f_0 2t_D); \\ &\dots \\ x(n) &= \sin(2\pi f_0 nt_D). \end{aligned} \tag{1.2}$$

Оскільки два значення синусоїди є однаковими, якщо їх значення аргументів відрізняються на 2π радіан, тобто

$$x(\psi) = \sin(\psi + 2\pi m),$$

де m – будь яке ціле число, то можна записати

$$x(n) = \sin(2\pi f_0 nt_D) = \sin(2\pi f_0 nt_D + 2\pi m) = \sin(2\pi(f_0 + m/nt_D)nt_D). \tag{1.3}$$

Якщо виберемо m кратним n , тобто $m = kn$, то маємо

$$x(n) = \sin(2\pi(f_0 + k/t_D)nt_D). \tag{1.4}$$

Підставивши $f_D = 1/t_D$ в (1.4) і прирівнявши його до (1.2), отримуємо

$$x(n) = \sin(2\pi f_0 nt_D) = \sin(2\pi(f_0 + kf_D)nt_D). \tag{1.5}$$

Отже множники f_0 і $(f_0 + kf_D)$ дають однакові результати. Це означає, що множина послідовних дискретних відліків $x(n)$, яка описує синусоїду з частотою f_0 , також точно описує синусоїди з іншими частотами $(f_0 + kf_D)$.

Іншими словами: при дискретизації з частотою f_D ми не зможемо відрізнити дискретні відліки синусоїди частотою f_0 Гц від відліків синусоїди частотою $(f_0 + kf_D)$ Гц, де k – будь яке ціле додатне чи від'ємне число.

Така ситуація спостерігається для спектру будь-якої дискретної послідовності – він містить періодично повторювані копії вихідного безперервного сигналу.

Період повторення копій в частотній області дорівнює f_D , а сам періодичний спектр має протяжність від постійної складової до

нескінченності в обох напрямках частотної осі (оскільки k може бути і додатним, і від'ємним числом).

Розглянемо дискретизацію синусоїдального сигналу частотою $f_0 = 7$ кГц з частотою $f_D = 6$ кГц. Відліки на рис. 1.14 позначено квадратними точками, за 7 періодів сигналу отримано 6 відліків. Ці відліки є абсолютно аналогічними відлікам, які б отримали при дискретизації синусоїди частотою 1 кГц. Тобто, тут синусоїда частотою 7 кГц виступає під псевдонімом (alias) синусоїди частотою 1 кГц.

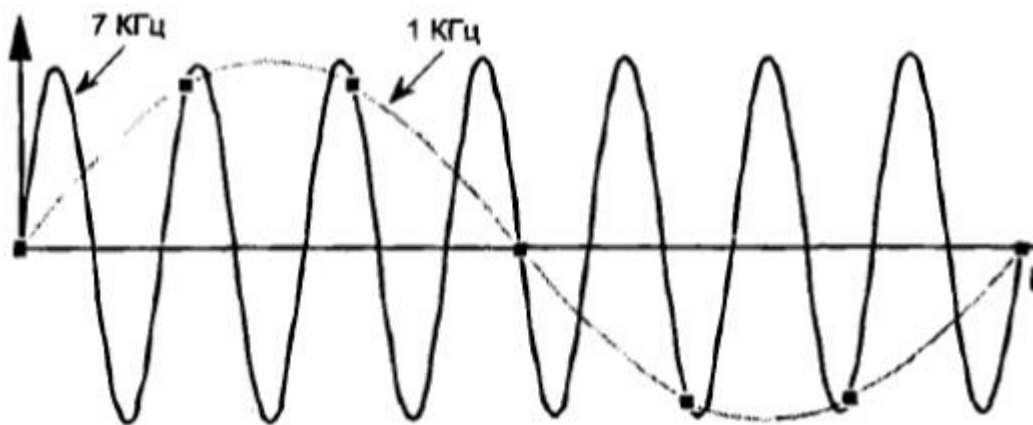


Рис. 1.14. Дискретизація синусоїди частотою 7 кГц

Параметри формули (1.5) тут наступні:

$$f_0 = 7 \text{ кГц}; f_D = 6 \text{ кГц}, k = -1; (f_0 + kf_D) = 7 - 6 = 1 \text{ кГц}.$$

Вибір частоти дискретизації виконується на основі теореми Котельникова-Шеннона:

Довільний сигнал $s(t)$, неперервний спектр якого не містить частот вище $f_{\text{гр}} = \omega_{\text{гр}}/2\pi$ може бути точно відновлений (інтерпольований) за його відліками, які узяті через рівні інтервали часу $t = k\Delta t = k\pi/\omega_{\text{гр}}$.

Тобто гранична частота сигналу (так звана *частота Найквіста*) дорівнює половині частоти дискретизації. З допомогою цієї теореми може бути вибраний оптимальний крок дискретизації реального сигналу і оцінена похибка дискретизації, що виникає при цьому.

Проте використання теореми як точного твердження по відношенню до реальних сигналів натрапляє на ряд принципових труднощів.

По-перше, реальний сигнал має кінцеву тривалість і, отже, має

необмежений спектр. Проте через реальні властивості джерел сигналів і обмеженість смуги пропускання реальних приладів і систем спектр сигналу з тією або іншою мірою точності можна вважати обмеженим деякою граничною частотою. Найчастіше граничне значення частоти $f_{\text{гр}} = \omega_{\text{гр}} / 2\pi$ визначають на основі енергетичного критерію, за яким практичну ширину спектру сигналу вибирають так, щоб в ній була зосереджена значна частина енергії сигналу.

Для цього використовують рівність Парсеваля, що дозволяє визначити енергію сигналу E_S або через функцію $s(t)$, що описує реальний сигнал тривалістю T , або через модуль її спектральної щільності $|\dot{S}(\omega)|$

$$E_s = \int_{-\infty}^{\infty} |s(t)|^2 dt = \int_0^T |s(t)|^2 dt = \frac{1}{\pi} \int_0^{\infty} |\dot{S}(\omega)|^2 d\omega.$$

Практична ширина спектру сигналу, яка зосереджена в діапазоні частот від 0 до деякого значення $\omega_{\text{гр}}$, визначається з співвідношення

$$\frac{1}{\pi} \int_0^{\omega_{\text{гр}}} |\dot{S}(\omega)|^2 d\omega = \eta \int_0^{\infty} |\dot{S}(\omega)|^2 d\omega, \quad (1.6)$$

де $\omega_{\text{гр}}$ - гранична частота, що визначає верхнє значення спектру сигналу; η – коефіцієнт досить близький до 1 (на практиці його значення залежно від вимог до якості відтворення сигналу вибирають в інтервалі від 0,9 до 0,998). Значення $\eta = 0,95$ означає, що в смузі частот від 0 до $\omega_{\text{гр}}$ міститься 95 % енергії сигналу. Значення граничної частоти знаходять, вирішуючи трансцендентне рівняння (1.6).

Обмеження спектру реального сигналу, природно, призводить до спотворення сигналу. Таким чином, відновлення обмеженого в часі сигналу по відліках відповідно до теореми Котельникова-Шеннона за умови примусового обмеження спектру сигналу можливо тільки приблизно. Точність такого наближення може бути оцінена як абсолютним значенням похибки, званою енергією похибки

$$\Delta E = \frac{1}{\pi} \int_{\omega_{\text{гр}}}^{\infty} |\dot{S}(\omega)|^2 d\omega,$$

так і відносною похибкою:

$$\gamma = \frac{\Delta E}{E_s}, \text{ де } E_s = \frac{1}{\pi_0} \int_0^\infty |\dot{S}(\omega)|^2 d\omega.$$

По-друге, також і за рахунок кінцевої кількості відліків на інтервалі тривалості сигналу $T=T_2-T_1$, яка відповідно до теореми Котельникова-Шеннона складе $2f_{гр}T + 1 \approx 2f_{гр}T$ (оскільки зазвичай $f_{гр}T \gg 1$). Ця складова є наслідком нехтування вкладом нескінченного числа функцій відліків, відповідних вибіркам за межами інтервалу (T_1, T_2) .

На практиці частоту дискретизації вибирають більшою, ніж $2f_{гр}$ із за того, що фільтр низьких частот, що використовується при відновленні сигналу, має не нескінченну крутизну спаду. Наприклад, діапазон звукових сигналів, що сприймаються людиною, не перевищує 20 кГц. Отже, при дискретизації записаних звукових сигналів з частотою не менше 40 кГц можна точно (для сприйняття людиною) відновити вихідний аналоговий сигнал за його цифровими відліками. Частота дискретизації звукового сигналу при запису на компакт-диск становить 44 100 Гц.

Зміст цієї теореми стане зрозумілим, якщо розглянути спектри сигналів $s(t)$ та $s_D(t)$:

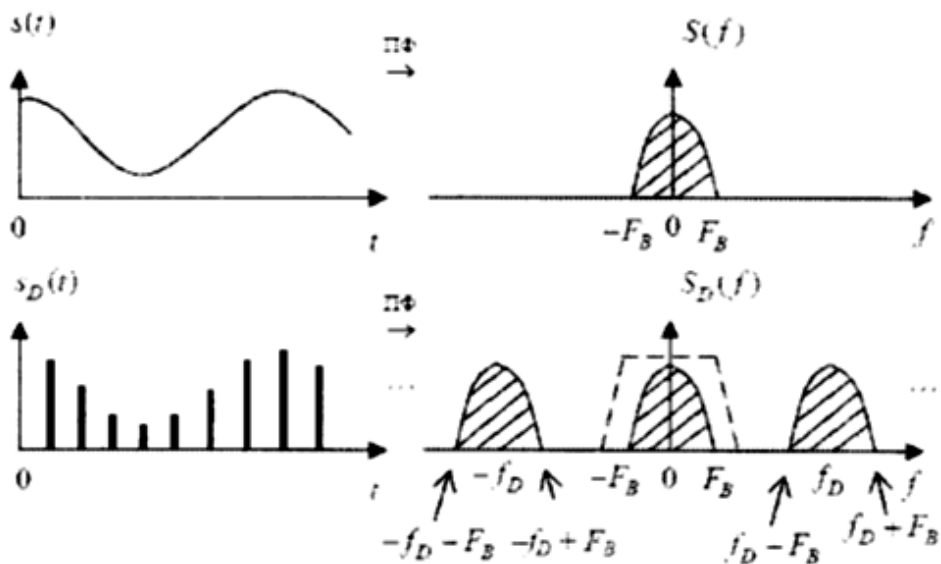


Рис.1.15. Теорема Котельникова

Спектр дискретного сигналу $S_D(f)$ містить в собі спектр аналогового сигналу $S(f)$ і нескінченну кількість копій $S(f)$, зсунутих на величину частоти дискретизації $f_D = f_D$. При $f_D > 2F_B$ копії не перетинаються з

основним пелюстком спектру $S_D(f)$ і точне відновлення сигналу $s(t)$ є можливим. При $f_D = 2F_B$ копії стикаються з основним пелюстком. Точне відновлення сигналу можливе тільки з допомогою ідеального фільтру низьких частот з нескінченною крутизною спаду АЧХ (на рисунку зображено АЧХ приблизно реального ФНЧ). При $f_D < 2F_B$ пелюстки перекриваються і точне відновлення сигналу неможливе.

Важливо щоб при дискретизації сигналу з частотою $2f_{гр}$ у ньому не виявилось складових з частотами вищими за $f_{гр}$ (*частота Найквіста*, яка дорівнює половині частоти дискретизації), оскільки це може викликати появу у спектрі відновленого після дискретизації сигналу складових, яких не було у вихідному сигналі. Це є наслідком ефекту, що має назву еліасинг (aliasing) – накладання або неможливість розрізнення різних неперервних сигналів при їхній дискретизації.

Припустимо, що при записі акустичного сигналу оркестру в приміщенні було присутнє джерело ультразвукового сигналу з частотою $f_1=32,05$ кГц. При запису з типовою частотою дискретизації сигналу на виході мікрофона $2f_{гр}=44,1$ кГц в динаміках при прослуховуванні такого запису з використанням ЦАП буде чути завадовий сигнал на частоті $2f_{гр}f_1 = f_{гр} - (f_1 - f_{гр}) = 44,1 - 32,05 = 12,05$ (кГц).

Як видно, в результаті дискретизації сигнали з частотами, що лежать вище половини частоти дискретизації, сприйматимуться з частотами «дзеркальними» відносно $f_{гр}$ в нижній частині спектру і складаються з присутніми там гармоніками.

Аналогічно в головний частотний діапазон дискретних сигналів «відображаються» й усі високочастотні шуми, присутні у вихідному аналоговому сигналі.

Таким чином, високочастотні складові накладаються на низькочастотні, в результаті цього відновлений аналоговий сигнал буде спотворюватись.

Для запобігання еліасингу слід підвищувати частоту дискретизації для вірного відображення усіх частотних складових сигналу або перед оцифруванням обмежувати спектр сигналу фільтрами низьких частот, які пропускають без зміни всі частоти, нижче заданої, і пригнічують в

сигналі частоти, вище заданої. Ця гранична частота називається частотою зрізу (cut-off frequency) фільтра. Частота зрізу антиеліасингових фільтрів встановлюється рівною половині частоти дискретизації. Такі ФНЧ майже завжди вбудовуються у реальні АЦП.

До задачі вибору частоти дискретизації сигналів повернемося пізніше після ознайомлення з спектральним аналізом обмежених у часі сигналів.

Питання для самоперевірки

1. Комплексні числа, форми подання.
2. Типи і причини виникнення завад.
3. Назвіть види тестових сигналів.
4. Наведіть класифікацію вимірювальних сигналів.
5. В чому полягає фільтрувальна властивість δ -функції?
6. Як виглядає узагальнена структура цифрових систем?
7. В чому суть аналого-цифрового перетворення?
8. В чому суть цифро-аналогового перетворення?
9. Назвіть основні оцінки точності відновлення (похибки дискретизації).
10. Сутність та причини появи ефекту еліасингу.
11. Що таке «частота Найквіста»?
12. В чому полягає фізичний сенс теореми Котельникова-Шеннона?

РОЗДІЛ 2. СПЕКТРАЛЬНИЙ АНАЛІЗ СИГНАЛІВ. ІНТЕГРАЛЬНІ ПЕРЕТВОРЕННЯ

2.1. Підходи до спектрального аналізу сигналів методом Фур'є

В теорії сигналів аналіз можна виконувати як в часовій (просторовій) області, так і в частотній області. Для цього необхідно отримати спектр сигналу (часовий або просторовий).

Поширеним у практиці спектрального аналізу як при аналоговій, так і при цифровій обробці є застосування ортогонального базису, утвореного гармонічними функціями з кратними частотами. Такий базис забезпечує точне розкладання сигналу і, крім цього, гармонічне коливання – єдине, що зберігає свою форму при проходженні через лінійні кола з постійними параметрами.

2.2. Розклад функції в ряд Фур'є

Розкладанню в ряд Фур'є піддаються періодичні сигнали. Вони представляються сумою гармонічних функцій або комплексних експонент з частотами, що утворюють арифметичну прогресію. Для того, щоб таке розкладання існувало, фрагмент сигналу $s(t)$ тривалістю в один період $[t_1, t_2]$, повинен задовольняти умовам Діріхле:

- інтервал $[t_1, t_2]$ можна розділити на кінцеву кількість інтервалів, на яких функція $s(t)$ є неперервною і монотонною, тобто число розривів першого роду (стрибків) має бути кінцевим;

- якщо t_0 є точкою розриву, то повинні існувати $s(t_0 + 0)$ і $s(t_0 - 0)$.

Ряд Фур'є може бути застосований для представлення не лише періодичних сигналів, але й сигналів кінцевої тривалості. При цьому обумовлюється часовий інтервал, для якого будується ряд Фур'є, а в інші моменти часу сигнал вважається рівним нулю. Для розрахунку коефіцієнтів ряду такий підхід фактично означає періодичне продовження сигналу за межами даного інтервалу.

Залежно від конкретної форми базисних функцій розглянемо три форми запису ряду Фур'є.

Перший випадок – *тригонометрична синус-косинусна форма*.

Системою ортонормованих базисних функцій на інтервалі періоду сигналу $[-T/2, T/2]$ у цьому випадку є

$$\begin{aligned}\varphi_0 &= \frac{1}{\sqrt{T}} \\ \varphi_1 &= \sqrt{2/T} \sin \omega_1 t, \\ \varphi_2 &= \sqrt{2/T} \cos \omega_1 t, \\ \varphi_3 &= \sqrt{2/T} \sin 2\omega_1 t, \\ \varphi_4 &= \sqrt{2/T} \cos 2\omega_1 t, \\ \varphi_5 &= \sqrt{2/T} \sin 3\omega_1 t, \\ \varphi_6 &= \sqrt{2/T} \cos 3\omega_1 t \dots\dots\dots,\end{aligned}$$

де $\omega_1 = 2\pi/T$.

Ряд Фур'є має вигляд

$$s(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} (a_i \cos i\omega_1 t + b_i \sin i\omega_1 t). \quad (2.1)$$

Гармонічне коливання з частотою $\omega_i = i\omega_1$ називається i -ю гармонікою сигналу. Коефіцієнти ряду a_i і b_i розраховуються по формулах

$$\begin{aligned}a_0 &= \frac{2}{T} \int_{-T/2}^{T/2} s(t) dt; \quad a_i = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \cos(i\omega_1 t) dt; \\ b_i &= \frac{2}{T} \int_{-T/2}^{T/2} s(t) \sin(i\omega_1 t) dt\end{aligned} \quad (2.2)$$

Таким чином, у загальному випадку періодичний сигнал є сумою постійної складової і нескінченного ряду гармонік з кратними частотами $\omega_i = i\omega_1$ ($i = 1, 2, \dots$).

Константа a_0 розраховується по загальній формулі для a_i як середнє значення сигналу на періоді:

$$\frac{a_0}{2} = \frac{1}{T} \int_{-T/2}^{T/2} s(t) dt.$$

Зауважимо, що границі інтегрування не обов'язково мають бути $[-T/2, T/2]$, як в формулах (2.2). Важливо, щоб інтервал інтегрування мав тривалість T – результат від цього принципово не зміниться. В окремих випадках може виявитись зручніше виконувати інтегрування, наприклад, від 0 до T або від $-T$ до 0.

Оскільки інтегрування ведеться в симетричних границях, то, якщо $s(t)$

- парна функція, то усі коефіцієнти b_i дорівнюватимуть нулю, і у формулі (2.1) будуть присутні тільки косинусні доданки. Якщо $s(t)$ є непарною функцією, дорівнювати нулю будуть, навпаки, косинусні коефіцієнти a_i і в формулі залишаться лише синусні доданки.

Другий випадок – *тригонометрична косинусна (дійсна) форма*.

Деяка незручність синус-косинусної форми ряду Фур'є полягає в тому, що для кожного значення індексу підсумовування i (тобто для кожної гармоніки з частотою $\omega_i = i\omega_1$) у формулі фігурують два доданки – синусний і косинусний. Використавши тригонометричне співвідношення

$$\cos(a - b) = \cos a \cdot \cos b + \sin a \cdot \sin b,$$

можна i -у гармоніку представити у вигляді

$$a_i \cos i\omega_1 t + b_i \sin i\omega_1 t = \sqrt{a_i^2 + b_i^2} \left(\frac{a_i}{\sqrt{a_i^2 + b_i^2}} \cos i\omega_1 t + \frac{b_i}{\sqrt{a_i^2 + b_i^2}} \sin i\omega_1 t \right) = A_i \cos(i\omega_1 t + \psi_i),$$

де $A_i = \sqrt{a_i^2 + b_i^2}$ – амплітуда і $\psi_i = -\arctg(b_i/a_i)$ – фаза.

Для знаходження початкової фази гармоніки $\psi_i = -\arctg(b_i/a_i)$ потрібно враховувати знаки коефіцієнтів a_i, b_i :

$$\psi_i = \begin{cases} -\arctg(b_i/a_i), & \text{якщо } a_i > 0; \\ \pi + \arctg(b_i/a_i), & \text{якщо } b_i > 0 \text{ і } a_i < 0; \\ -\pi + \arctg(b_i/a_i), & \text{якщо } b_i < 0 \text{ і } a_i < 0. \end{cases}$$

Це дозволяє трансформувати ряд (2.1) в іншу еквівалентну форму.

$$s(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} A_i \cos(i\omega_1 t + \psi_i). \quad (2.3)$$

Ця форма застосовується для спектрального представлення періодичного сигналу у вигляді двох залежностей амплітуд A_i і фаз ψ_i гармонік від частоти $\omega_i = i\omega_1$.

Якщо $s(t)$ є парною функцією, фази ψ_i можуть набувати тільки значень 0 і π (оскільки $b_i = 0$), а якщо $s(t)$ є непарною функцією, то можливі значення

для фази $-\pm \pi/2$ (оскільки $a_i = 0$).

Третій випадок – *комплексно-експоненційна форма*.

Ця форма представлення ряду Фур'є виходить з представлення косинуса у вигляді напівсуми комплексних експонент $\cos x = (e^{jx} + e^{-jx})/2$ та $\sin x = (e^{jx} - e^{-jx})/2j$, яке витікає з формули Ейлера $e^{jx} = \cos x + j \sin x$.

Застосувавши це перетворення до косинусної форми ряду Фур'є, отримаємо суми комплексних експонент з додатними і від'ємними показниками степені

$$s(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} \frac{A_i}{2} [\exp(ji\omega_1 t + j\psi_i) + \exp(-ji\omega_1 t - j\psi_i)].$$

Далі будемо розглядати експоненти зі знаком «мінус» в показнику як члени ряду з від'ємними номерами (тобто підсумовування почнеться не з одиниці, а з $-\infty$). Постійний доданок $a_0/2$ при цьому стане членом ряду з нульовим номером. В результаті матимемо комплексну форму запису ряду Фур'є

$$s(t) = \frac{a_0}{2} + \sum_{i=-\infty}^{\infty} \dot{C}_i \exp(ji\omega_1 t). \quad (2.4)$$

Комплексні коефіцієнти ряду (2.4) пов'язані з амплітудами A_i і фазами ψ_i , які є в дійсній формі запису ряду Фур'є, наступними співвідношеннями:

$$\dot{C}_i = \frac{1}{2} A_i \exp(j\psi_i); \quad A_i = 2|\dot{C}_i|; \quad \psi_i = \arg(\dot{C}_i).$$

Оскільки $A_i = \sqrt{a_i^2 + b_i^2}$ і $\psi_i = -\arctg(b_i/a_i)$, то комплексні коефіцієнти зв'язані з коефіцієнтами a_i та b_i синус-косинусної форми ряду Фур'є як:

$$\dot{C}_i = \frac{1}{2} (a_i - jb_i); \quad a_i = 2\operatorname{Re}(\dot{C}_i); \quad b_i = 2\operatorname{Im}(\dot{C}_i).$$

Звідси отримаємо формулу для безпосереднього розрахунку коефіцієнтів \dot{C}_i ряду Фур'є в комплексній формі (оскільки $e^{-jx} = \cos x - j \sin x$)

$$\dot{C}_i = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \frac{\cos i\omega_1 t + j \sin i\omega_1 t}{2} dt = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \exp(-ji\omega_1 t) dt. \quad (2.5)$$

Система ортонормованих базисних функцій в цьому випадку має вигляд

$$\{\varphi_i\} = \exp(-j\omega_1 t)/\sqrt{T}, i=0, \pm 1, \pm 2, \pm 3 \dots$$

Якщо $s(t)$ є парною функцією, коефіцієнти ряду \dot{C}_i будуть дійсними, а якщо $s(t)$ є непарною функцією, то коефіцієнти ряду виявляться чисто уявними.

Сукупність амплітуд гармонік ряду Фур'є часто називають *амплітудним спектром*, а сукупність їх фаз – *фазовим спектром* (ці поняття не слід плутати з амплітудною і фазочастотною характеристиками, які відносяться не до сигналів, а до ланцюгів).

Якщо аналізований сигнал $s(t)$ є дійсним, то його амплітудний і фазовий спектри мають симетрію:

$$\dot{C}_{-i} = \dot{C}_i^*, |\dot{C}_{-i}| = |\dot{C}_i^*|, \psi_{-i} = -\psi_i.$$

Узагальнимо *властивості розкладання в ряд Фур'є періодичних сигналів* на прикладі комплексно-експоненційного ряду Фур'є:

1. Спектр періодичного сигналу повністю визначають *дві характеристики* – *амплітудна і фазова*, тобто, модулі та аргументи комплексних коефіцієнтів ряду Фур'є.

2. Спектр періодичного сигналу є *дискретним* або *лінійчастим*. Він складається з окремих ліній (точок) на відповідних частотах $\dots, -2\omega_1, -\omega_1, 0, \omega_1, 2\omega_1, \dots$.

3. Спектри періодичних сигналів є *еквідистантними*, тобто сусідні спектральні лінії знаходяться одна від одної на однаковій відстані $\omega_1 = 2\pi/T$ ($f_1 = \frac{1}{T}$ – в лінійних частотах). Частота ω_1 називається частотою першої гармоніки.

4. Коефіцієнти ряду Фур'є, а отже й спектри є *симетричними* відносно частоти $\omega=0$. Оскільки

$$\dot{C}_i = \frac{1}{T} \int_{-T/2}^{T/2} s(t) e^{-ji\omega_1 t} dt = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \cos i\omega_1 t dt - j \frac{1}{T} \int_{-T/2}^{T/2} s(t) \sin i\omega_1 t dt,$$

то, якщо сигнал $s(t)$ є дійсною функцією часу, перша складова правої частини коефіцієнта \dot{C}_i є дійсною, а друга складова – уявною, і тоді можна записати

$$\dot{C}_i = \begin{cases} (a_i - jb_i)/2, & i > 0, \\ (a_i + jb_i)/2, & i < 0. \end{cases}$$

З цього виразу виходить, що $\dot{C}_{-i} = \dot{C}_i^*$. Оскільки $\dot{C}_i = \frac{1}{2}A_i \exp(j\psi_i)$, то отримаємо, що $|\dot{C}_{-i}| = |\dot{C}_i^*|$, $\psi_{-i} = -\psi_i$.

Таким чином, якщо сигнал $s(t)$ є дійсною функцією часу, то модулі коефіцієнтів ряду Фур'є (амплітудний спектр) мають властивість парної симетрії (відображаються дзеркально відносно вертикальної осі) відносно i (тобто частоти), а аргументи (фазовий спектр) – властивість непарної симетрії (відображаються дзеркально відносно початку системи координат) відносно i .

5. Якщо $s(t)$ є дійсною й парною функцією часу, то уявна частина коефіцієнтів ряду Фур'є \dot{C}_i дорівнює нулю, тобто коефіцієнти синус-косинусного ряду Фур'є $b_i = 0$. З умови $\operatorname{tg}\psi_i = -b_i/a_i$ виходить, що ψ_i може набувати лише двох значень: 0 або $\pm\pi$.

Якщо $s(t)$ є дійсною і непарною функцією часу, то дійсна частина коефіцієнтів ряду Фур'є \dot{C}_i дорівнює нулю, тобто коефіцієнти синус-косинусного ряду Фур'є $a_i = 0$. У цьому випадку ψ_i може набувати лише двох значень: $\pi/2$ або $-\pi/2$.

Розглянемо найпростіший приклад розкладання в ряд Фур'є для послідовності прямокутних імпульсів з амплітудою A , тривалістю імпульсу τ і періодом повторення T . Початок відліку часу приймемо розташованим в середині імпульсу (рис. 2.1). Сигнал доцільно описати з використанням функції Хевісайда $\sigma(t)$

$$s(t) = \sum_{i=-\infty}^{\infty} A[\sigma(t + \frac{\tau}{2} - kT) - \sigma(t - \frac{\tau}{2} - kT)].$$

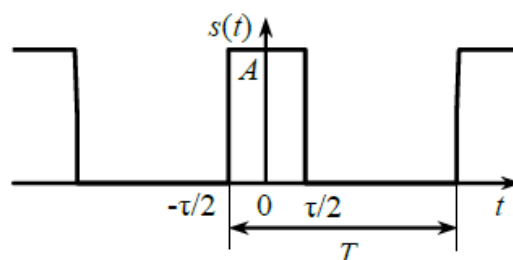


Рис. 2.1. Періодична послідовність прямокутних імпульсів

Цей сигнал є парною функцією, тому для його представлення зручніше використовувати синус-косинусну форму ряду Фур'є – в ній будуть присутні тільки косинусні доданки

$$a_i = \frac{2}{T} \int_{-\tau/2}^{\tau/2} A \cos\left(\frac{2\pi i}{T} t\right) dt = \frac{2A}{\pi i} \sin\left(\frac{\pi i \tau}{T}\right).$$

В отриманій формулі можна помітити, що тривалість імпульсів і їх період слідування входять в неї виключно у вигляді відношення періоду до тривалості імпульсів, яке називають *шпаруватістю* (прогальністю). Шпаруватість завжди більше одиниці і дає змогу оцінювати можливість накопичення великих енергій під час пауз і генерування цієї енергії протягом короткочасної тривалості імпульсу.

Величина, зворотна шпаруватості, називається *коефіцієнтом заповнення*. Якщо позначити шпаруватість як $q=T/\tau$, то спектр Фур'є послідовності прямокутних імпульсів набуде вигляду

$$s(t) = \frac{A}{q} + \sum_{i=1}^{\infty} \frac{2A}{q} \frac{\sin\left(\frac{\pi i}{q}\right)}{\frac{\pi i}{q}} \cos \frac{2\pi i}{T} t = \frac{A}{q} + \sum_{i=1}^{\infty} \frac{2A}{\pi i} \sin\left(\frac{\pi i}{q}\right) \cos \frac{2\pi i}{T} t.$$

Амплітуди гармонічних доданків ряду залежать від номера гармоніки по закону $\sin(x)/x$ (рис. 2.2). Графік функції $\sin(x)/x$ має пелюстковий характер. Говорячи про ширину цих пелюсток, слід підкреслити, що для графіків дискретних спектрів періодичних сигналів можливі два варіанти градуювання горизонтальної осі – в номерах гармонік і в частотах [6].

На рис. 2.2, а) градуювання осі абсцис виконане за частотою, а на рис. 2.2, б) - у комплексному представленні градуювання відповідає номерам гармонік, частотні ж параметри спектру нанесені на графік за допомогою розмірних ліній. Амплітудний спектр у дійсній формі має завжди додатні коефіцієнти ($i \geq 0$), а в комплексній формі – i може бути будь-яким цілим числом від $-\infty$ до $+\infty$.

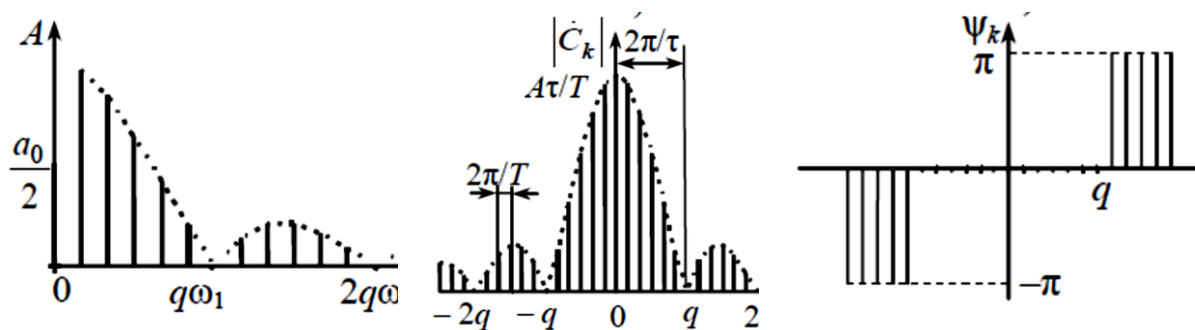


Рис. 2.2. Спектр періодичної послідовності прямокутних імпульсів:
 а) амплітудний дійсна форма, б) амплітудний комплексна форма;
 в) фазовий спектр

Отже, ширина пелюсток, виміряна у кількості гармонік, дорівнює шпаруватості послідовності. При $i = q$ маємо $\sin\left(\frac{\pi i}{q}\right) = 0$. Звідси витікає важлива властивість спектру послідовності прямокутних імпульсів – в ньому відсутні (мають нульові амплітуди) гармоніки з номерами, кратними шпаруватості.

Відстань по частоті між сусідніми гармоніками дорівнює частоті слідування імпульсів $2\pi/T$ (в лінійних частотах – $1/T$). Ширина пелюсток спектру, виміряна в одиницях частоти, становить $2\pi/\tau$, тобто є обернено пропорційною тривалості імпульсів. Це є властивістю спектру Фур'є: чим коротшим є сигнал, тим ширшим є його спектр.

2.3. Інтегральне перетворення Фур'є

Розглянутий метод розкладення періодичних сигналів в ряд Фур'є може бути узагальнений і на випадок неперіодичних сигналів, які відповідають умовам Діріхле і є абсолютно інтегрованими на усій числовій осі $\int_{-\infty}^{\infty} |s(t)| dt < \infty$. При цьому неперіодична функція розглядається як граничний випадок періодичної функції при $T \rightarrow \infty$.

Для отримання спектру неперіодичного сигналу зручніше використовувати комплексну форму ряду Фур'є (2.4). Нехай аналізований сигнал $s(t)$ є одиночним (імпульсним) сигналом обмеженої тривалості. Якщо уявно доповнити цей сигнал послідовністю таких самих сигналів з періодом T , то такий сигнал може бути представлений рядом Фур'є

$$s_{\text{пер}}(t) = \sum_{i=-\infty}^{\infty} \dot{C}_i \exp(ji\omega_1 t) \quad (2.6)$$

з коефіцієнтами:

$$\dot{C}_i = \frac{1}{T} \int_{-T/2}^{T/2} s(t) \exp(-ji\omega_1 t) dt. \quad (2.7)$$

Якщо спрямувати період T цієї послідовності до нескінченності, то матимемо:

1. Частоти сусідніх гармонік $i\omega_1$ та $(i-1)\omega_1$ будуть скільки завгодно близькими, тому у формулах (2.6) і (2.7) дискретну змінну $i\omega_1$ можна змінити на неперервну частоту ω , а значення частоти $\omega_1 \rightarrow d\omega$.

2. Амплітудні коефіцієнти \dot{C}_i стануть нескінченно малими, оскільки у знаменнику (2.7) є нескінченно велике значення періоду T .

Підставимо коефіцієнти \dot{C}_i в (2.6), враховуючи, що $\frac{1}{T} = \frac{\omega_1}{2\pi}$.

$$s(t) = \lim_{T \rightarrow \infty} s_{\text{пер}}(t) = \lim_{T \rightarrow \infty} \sum_{i=-\infty}^{\infty} \frac{\omega_1}{2\pi} \left[\int_{-\frac{T}{2}}^{\frac{T}{2}} s(t) \exp(-j\omega_1 t) dt \right] \exp(j\omega_1 t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(j\omega t) \left[\int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt \right] d\omega. \quad (2.8)$$

У подвійному інтегралі внутрішній інтеграл називається *спектральною щільністю* (або просто – *спектром*) $\dot{S}(\omega)$ сигналу $s(t)$ і є *перетворенням Фур'є* від функції сигналу. Тут границі інтегрування змінилися, оскільки при $T \rightarrow \infty$ також і $T/2 \rightarrow \infty$.

Із (2.8) витікає, що відновлення сигналу за його спектральною щільністю можливе як

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{S}(\omega) \exp(j\omega t) d\omega. \quad (2.9)$$

Формула (2.9) описує обернене перетворення Фур'є

Якщо використовувати не кругову частоту ω , а лінійну частоту f , формули прямого і оберненого перетворення Фур'є стають більш симетричними, відрізняючись лише знаком у показнику експоненти:

$$\begin{aligned} \dot{S}(f) &= \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt; \\ s(t) &= \int_{-\infty}^{\infty} \dot{S}(f) \exp(j2\pi ft) df. \end{aligned}$$

Якщо аналізований сигнал $s(t)$ є дійсною функцією, то відповідна йому спектральна щільність $\dot{S}(\omega)$ є «спряжено-симетричною» відносно нульової частоти. Це означає, що значення спектральної щільності на частотах ω і $-\omega$ є комплексно-спряженими по відношенню одне до одного $\dot{S}(-\omega) = \dot{S}^*(\omega)$.

Якщо $s(t)$ є парною функцією, то, як і у разі ряду Фур'є, спектр буде чисто дійсним (і, отже, буде парною функцією). Якщо, навпаки, $s(t)$ – функція непарна, то спектральна функція $\dot{S}(\omega)$ буде чисто уявною (і непарною). Модуль спектральної щільності часто називають *амплітудним спектром*, а її аргумент – *фазовим спектром*. Для дійсного сигналу амплітудний спектр є парною функцією частоти, а фазовий – непарною:

$$|\dot{S}(-\omega)| = |\dot{S}(\omega)|, \quad \psi(-\omega) = -\psi(\omega).$$

Отже, перетворення Фур'є ставить у відповідність сигналу, заданому в часі, його спектральну функцію (щільність). При цьому здійснюється перехід з часової області в частотну. Перетворення Фур'є є взаємно-однозначним, тому представлення сигналу в частотній області (спектральна функція) містить рівно стільки ж інформації, скільки й вихідний сигнал, заданий в часовій області.

Відмінність ряду і перетворення Фур'є полягає в тому, що:

- ряд Фур'є подає періодичну функцію у вигляді нескінченної кількості гармонічних складових з дискретними частотами;
- перетворення Фур'є подає неперіодичну функцію у вигляді нескінченної кількості гармонічних складових з неперервною послідовністю частот.

Властивості перетворення Фур'є.

Під властивостями перетворення Фур'є розуміється взаємна відповідність трансформацій сигналів і їх спектрів. Добре знання властивостей перетворення Фур'є дозволяє передбачати приблизний (а іноді і точний) вид спектру аналізованого сигналу і таким чином контролювати правдоподібність результату при обчисленнях на комп'ютері.

Розглянемо два сигнали $f(t)$ і $g(t)$, які мають спектральні щільності відповідно $F(\omega)$ і $G(\omega)$.

1. *Лінійність*

Перетворення Фур'є є лінійним інтегральним перетворенням. Сенс властивості лінійності можна сформулювати так: спектр суми дорівнює сумі спектрів.

Говорячи математичною мовою, лінійна комбінація сигналів має спектр у вигляді такої ж (з тими ж коефіцієнтами) лінійної комбінації їх спектральних щільностей:

$$\text{Якщо } s(t) = \alpha f(t) + \beta g(t), \text{ то } \hat{S}(\omega) = \alpha \hat{F}(\omega) + \beta \hat{G}(\omega).$$

2. *Затримка*

Розглянемо, як впливає на спектральну функцію затримка сигналу в часі. Нехай τ – час затримки, тоді сигнал $s(t) = f(t - \tau)$. Його спектральна функція зміниться таким чином:

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} f(t-\tau) e^{-j\omega t} dt = e^{-j\omega t} \int_{-\infty}^{\infty} f(t-\tau) e^{j\omega(t-\tau)} d(t-\tau) = \dot{F}(\omega) e^{-j\omega t}.$$

Результат показує, що спектр незатриманого сигналу $f(t)$ виявився помноженим на комплексну експоненту виду $e^{-i\omega t}$. Таким чином, амплітудний спектр сигналу не змінюється (оскільки модуль такої комплексної експоненти дорівнює 1. Здоровий глузд підказує, що співвідношення між амплітудами спектральних складових із-за зсуву сигналу в часі змінитися не повинно). Фазовий спектр набуває додаткового доданку $-\omega\tau$, який лінійно залежить від частоти.

3. Зміна масштабу осі часу

Розглядаючи конкретні приклади, ми вже вказали загальне на практиці правило: чим коротше сигнал, тим ширше його спектр. Тепер поглянемо та це правило з боку строгих теоретичних позицій. Якщо змінити тривалість сигналу $f(t)$, зберігаючи його форму, то новий сигнал $s(t)$ слід записати як $s(t) = f(at)$.

При $a > 1$ сигнал стискується, при $a < 1$ – розтягується, якщо $a < 0$, додатково відбувається дзеркальне відображення сигналу відносно вертикальної осі. Подивимось, як таке перетворення позначається на спектрі:

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} f(at) e^{-j\omega t} dt = \frac{1}{a} \int_{-\infty}^{\infty} f(x) e^{-j\frac{\omega}{a}x} dx = \frac{1}{a} \dot{F}\left(\frac{\omega}{a}\right).$$

Отже, зміна тривалості сигналу веде до зміни ширини спектру в протилежний бік (аргумент t множиться на a , а ω – ділиться) у сполученні із збільшенням (при розтягуванні, $a < 1$) або зменшенням (при стисканні, $a > 1$) рівня спектральних складових.

Отримана формула є справедливою для $a > 0$. При $a < 0$ використана заміна змінної $a\tau = x$ викличе перестановку границь інтегрування і, як наслідок, зміну знаку у результаті:

$$\dot{S}(\omega) = \frac{1}{a} \dot{F}\left(\frac{\omega}{a}\right), a < 0.$$

Об'єднуючи обидва випадки, отримаємо

$$\dot{S}(\omega) = \frac{1}{|a|} \dot{F}\left(\frac{\omega}{a}\right), a \neq 0.$$

У окремому випадку $a = -1$ формула має вигляд

$$\dot{S}(\omega) = \dot{F}(-\omega) = \dot{F}^*(\omega).$$

Отже, дзеркальне відображення сигналу відносно початку відліку часу призводить до дзеркального відображення спектру відносно нульової частоти. Для дійсного сигналу це відповідає комплексному сполученню спектру.

В даному випадку результат не отримується множенням вихідного спектру на деяку функцію. Це означає, що зміна тривалості сигналу не може бути здійснена лінійною системою з постійними параметрами.

4. Спектр згортки сигналів

Згортка сигналів є дуже часто використовуваною в вимірювальній техніці інтегральною операцією. Вона описує, зокрема, проходження сигналу $s(t)$ через лінійну систему з постійними параметрами

$$s(t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau.$$

де функцію $g(t-\tau)$ називають *ядром згортки* (kernel) або імпульсною характеристикою лінійної системи.

Піддамо таку конструкцію перетворенню Фур'є:

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\tau) g(t-\tau) dt e^{-j\omega t} dt = \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} d\tau \int_{-\infty}^{\infty} g(t-\tau) e^{-j\omega(t-\tau)} d(t-\tau) = \dot{F}(\omega) \dot{G}(\omega).$$

Отриманий результат дуже важливий, він часто використовується на практиці: *спектр згортки сигналів дорівнює добутку спектрів*.

У випадку, коли сигнал є добутком двох інших $s(t) = f(t) \cdot g(t)$, можна кожну з цих функцій представити їх оберненими перетвореннями Фур'є від відповідних спектральних щільностей кожного з сигналів

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{F}(\omega) \exp(j\omega t) d\omega,$$

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{G}(\omega) \exp(j\omega t) d\omega$$

і отримати наступне співвідношення

$$\dot{S}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) \left[\int_{-\infty}^{\infty} \dot{G}(x) \exp(jxt) dx \right] e^{-j\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{G}(x) \left[\int_{-\infty}^{\infty} f(t) e^{-j(\omega-x)t} dt \right] dx.$$

Інтеграл у квадратних дужках по змінній t є спектральною щільністю функції $f(t)$ при частоті $(\omega-x)$, тобто $\dot{F}(\omega-x)$, звідки маємо

$$\dot{S}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{G}(x) \dot{F}(\omega - x) dx$$

Таким чином, спектр добутку двох функцій часу $f(t) \cdot g(t)$ дорівнює (з коефіцієнтом $1/2\pi$) згортці їх спектрів $\dot{F}(\omega)$ і $\dot{G}(\omega)$, де

$$\int_{-\infty}^{\infty} \dot{G}(x) \dot{F}(\omega - x) dx = \dot{G}(x) * \dot{F}(\omega)$$

або інакше

$$f(t)g(t) \Leftrightarrow \frac{1}{2\pi} \dot{G}(x) * \dot{F}(\omega).$$

Неважко переконатися, що операція згортки є комутативною

$$\dot{G}(\omega) * \dot{F}(\omega) = \dot{F}(\omega) * \dot{G}(\omega).$$

5. Рівність Парсеваля

Якщо сигнали $f(t)$ і $g(t)$ будуть однаковими, то з попереднього рівняння отримаємо рівність

$$\int_{-\infty}^{\infty} f(t)g(t) dt = \int_{-\infty}^{\infty} f^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{F}(x) \dot{F}(\omega - x) dx,$$

яка у випадку, коли $\omega=0$, набуде вигляду

$$\int_{-\infty}^{\infty} f^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{F}(x) \dot{F}(-x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{F}(\omega) \dot{F}^*(\omega) d\omega.$$

Остаточно, можна записати рівність Парсеваля для неперіодичного сигналу у вигляді

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\dot{F}(\omega)|^2 d\omega = E$$

Ця рівність встановлює зв'язок між енергією сигналу E та його спектральною щільністю. Залежність $|\dot{F}(\omega)|^2$ при цьому має розмірність енергії, що припадає на 1 Гц, і може трактуватись як спектральна щільність енергії сигналу.

Приклади обчислення спектрів.

Обчислимо спектр *прямокутного імпульсу*, центрованого відносно початку відліку часу (рис. 2.3):

$$s(t) = A \left[\sigma \left(t + \frac{\tau}{2} \right) - \left(t - \frac{\tau}{2} \right) \right].$$

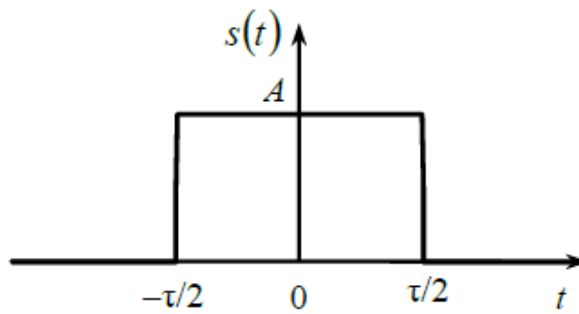


Рис. 2.3. Прямокутний імпульс

$$\begin{aligned} \dot{S}(\omega) &= \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt = A \int_{-\tau/2}^{\tau/2} \exp(-j\omega t) dt = A \int_{-\tau/2}^{\tau/2} (\cos\omega t - \\ &- j\sin\omega t) dt = A \int_{-\tau/2}^{\tau/2} \cos\omega t dt = \frac{2A}{\omega} \sin \frac{\omega\tau}{2} = A\tau \frac{\sin \frac{\omega\tau}{2}}{\frac{\omega\tau}{2}}. \end{aligned} \quad (2.10)$$

Вид спектру сигналу зображено на рис. 2.4.

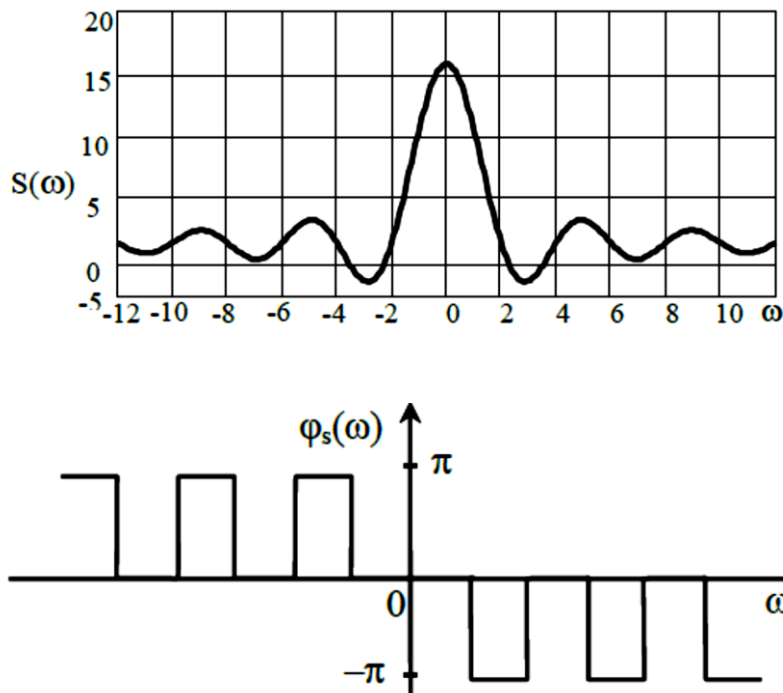


Рис. 2.4. Спектральна щільність і фазовий спектр прямокутного імпульсу з параметрами: $\tau = \pi$ с; $A = 5$ В

Амплітудний спектр має пелюстковий характер, ширина пелюсток складає $2\pi/\tau$, тобто є обернено пропорційною тривалості імпульсу. Значення спектральної щільності на нульовій частоті дорівнює площі імпульсу $S(0)=A\tau$ (на графіку $A\tau \approx 15,7$).

Спектральна функція є дійсною, тому фазовий спектр набуває лише два

значення -0 або π залежно від знаку функції $\sin(x)/x$. Значення фази π і $-\pi$ однакові, різні знаки для фазового спектру при $\omega > 0$ і $\omega < 0$ використані з метою представлення у вигляді непарної функції згідно з його загальними властивостями.

Розглянемо випадок, коли імпульс зсувається в часі. Нехай імпульс починається в нульовий момент часу (рис. 2.5).

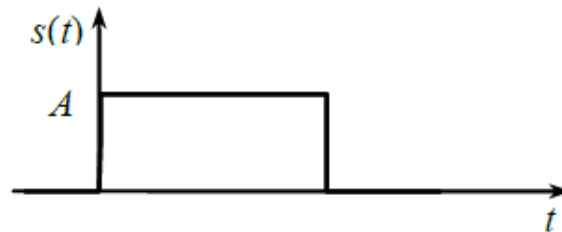


Рис. 2.5. Прямокутний імпульс із затримкою в часі на $\tau/2$

Обчислимо його спектральну функцію з врахуванням формули Ейлера :

$$\sin x = \frac{1}{2j}(e^{jx} - e^{-jx})$$

$$\begin{aligned} \dot{S}(\omega) &= A \int_0^{\tau} \exp(-j\omega t) dt = \frac{A}{j\omega} \int_{-j\omega\tau}^0 \exp(z) dz = \frac{A}{j\omega} [\exp(0) - \\ &\exp(-j\omega\tau)] = \frac{A}{j\omega\tau} [\exp(j\omega\tau/2) - \exp(-j\omega\tau/2)] \exp(-j\omega\tau/2) = \\ &A\tau \frac{\sin(\omega\tau/2)}{\omega\tau/2} \exp(-j\omega\tau/2). \end{aligned} \quad (2.11)$$

Спектри сигналу зображено на рис. 2.6 [4].

З формул (2.10) і (2.11) і графіків (рис. 2.4 і 2.5) видно, що після затримки імпульсу в часі його амплітудний спектр залишився тим самим, а фазовий отримав лінійно залежний від частоти зсув.

Цей приклад демонструє прояв властивості перетворення Фур'є, що стосується зміни спектру при зсуві сигналу в часі.

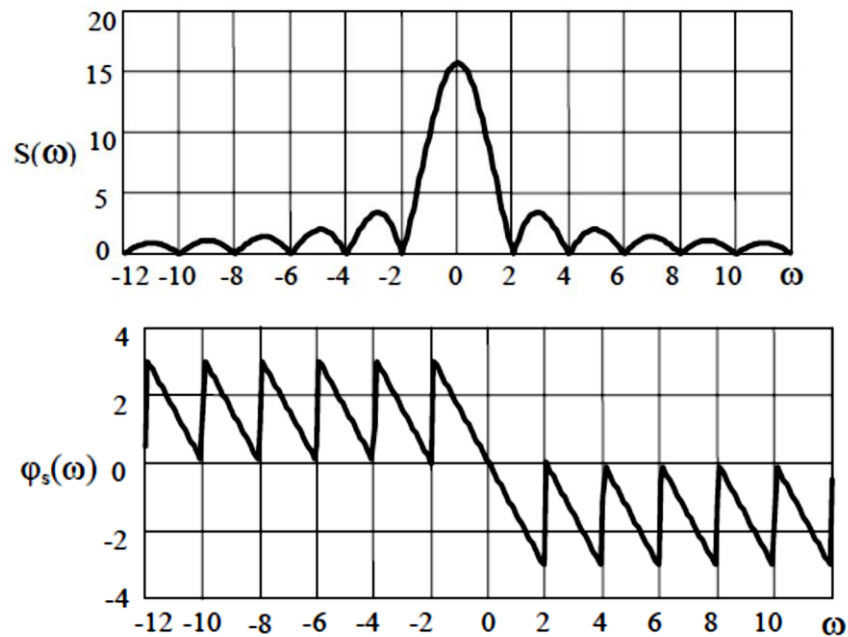


Рис. 2.6. Амплітудний і фазовий спектри прямокутного імпульсу із затримкою

Спектр цього сигналу тягнеться до нескінченності, лише поступово затухаючи. Тому вводять поняття ефективної ширини спектру. Як видно з графіків, спектр має пелюстковий характер і ширина головної пелюстки дорівнює $2\pi/\tau$. Саме її приймають за ефективну ширину спектру, яка є обернено пропорційною тривалості імпульсу. Це загальне співвідношення: чим коротшим є сигнал, тим ширшим є його спектр. Добуток же ефективних значень тривалості сигналу і ширини його спектру B (база сигналу) залишається рівним деякій константі, залежній від конкретного способу визначення цих параметрів. У розглянутому прикладі B , очевидно, дорівнює 2π . Взагалі, для сигналів простої форми (що не мають складної внутрішньої імпульсної структури) величина бази незалежно від способу визначення ефективних значень тривалості і ширини спектру складає декілька одиниць.

Тривалість сигналу і ширина його спектру підкоряються співвідношенню невизначеності, тобто добуток цих параметрів (база сигналу) не може бути меншим за одиницю. Обмежень максимального значення бази сигналу не існує. Звідси витікає, що можна сформувати сигнал великої тривалості, що одночасно має і широкий спектр (такі

сигнали називають широкосмуговими, або складними, або сигналами з великою базою). А ось короткий сигнал з вузьким спектром, згідно із співвідношенням невизначеності, існувати не може.

При введенні поняття перетворення Фур'є були вказані умови його застосовності: виконання умов Діріхле і абсолютна інтегрованість сигналу. Проте у ряді випадків можна застосувати перетворення Фур'є і до сигналів, що цим умовам не задовольняють, і отримати при цьому цілком осмислений і практично корисний результат. скористаємося перетворенням Фур'є для спектрального аналізу таких сигналів, до яких воно формально непридатне.

Дельта-функція.

Обчислимо перетворення Фур'є для сигналу у вигляді δ -функції. Виходячи з її фільтрувальної властивості маємо

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} \delta(t - t_0) \exp[-j\omega t] dt = \exp[-j\omega t_0].$$

В окремому випадку, коли $t_0 = 0$, маємо $\dot{S}(\omega) = 1$.

Спектр дельта-функції є константою, тобто є рівномірним в нескінченній смузі частот.

Це цілком узгоджується із загальним співвідношенням між тривалістю сигналу і шириною його спектру: дельта-імпульс має нескінченно малу тривалість, а його спектр є нескінченно широким. З отриманого результату виходить, що δ -функцію можна записати у вигляді оберненого перетворення Фур'є таким чином

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} d\omega.$$

Постійний за часом сигнал (константа).

Оскільки ми вже знаємо, що спектром дельта-функції є константа, завдяки дуальності перетворення Фур'є, можна відразу стверджувати, що спектром постійного сигналу $s(t)=A$ буде δ -функція частоти. Перевіримо це, скориставшись отриманим вище співвідношенням:

$$\dot{S}(\omega) = \int_{-\infty}^{\infty} A e^{-j\omega t} dt = 2\pi A \delta(\omega).$$

Знову добре простежується обернена пропорційність між тривалістю сигналу і шириною його спектру: нескінченно протяжний сигнал має

нескінченно вузький спектр.

Гармонічний сигнал.

Розрахуємо спектр гармонічного сигналу

$$s(t) = A \cos(\omega_0 t + \varphi_0).$$

Для розрахунку спектральної функції представимо косинус у вигляді півсуми комплексних експонент за формулою Ейлера і скористаємось формулою $\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} d\omega$:

$$\begin{aligned} \dot{S}(\omega) &= \int_{-\infty}^{\infty} A \cos(\omega_0 t + \varphi_0) e^{-j\omega t} dt = \frac{1}{2} \int_{-\infty}^{\infty} A e^{j\varphi_0} e^{-j(\omega - \omega_0)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} A e^{-j\varphi_0} e^{-j(\omega + \omega_0)t} dt = \\ &= A \pi e^{j\varphi_0} \delta(\omega - \omega_0) + A \pi e^{-j\varphi_0} \delta(\omega + \omega_0). \end{aligned}$$

Результатом є пара дельта-функцій, розташованих на частотах $\pm\omega_0$. Множники при них відповідають амплітуді і початковій фазі (тобто комплексній амплітуді) гармонічного сигналу.

Довільний періодичний сигнал.

Раніше показано, що періодичний сигнал з періодом T може бути представлений у вигляді ряду Фур'є

$$S(t) = \sum_{i=-\infty}^{\infty} \dot{C}_i \exp\left(j \frac{2\pi i}{T} t\right).$$

Після обчислення спектрів гармонічного сигналу і комплексної експоненти стає ясно, що спектральна функція такого сигналу є набором дельта-функцій, розташованих на частотах гармонік ряду Фур'є

$$\dot{S}(\omega) = \sum_{i=-\infty}^{\infty} 2\pi \dot{C}_i \delta\left(\omega - \frac{2\pi i}{T}\right).$$

Множники при дельта-функціях дорівнюють відповідним коефіцієнтам ряду Фур'є C_i , помноженим на 2π .

2.4. Дискретне перетворення Фур'є

За класифікацією, яка була наведена раніше, розглядалися відмінності аналогових і дискретних сигналів. Математична модель аналогового сигналу відповідає неперервній або кусково-неперервній функції, а дискретний сигнал $\mathbf{x}_d(t)$ є послідовністю відлікових значень $\{\dots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\}$ сигналу $\mathbf{x}(t)$ у відповідні моменти часу

$\{\dots, t_{-1}, t_0, t_1, t_2, \dots\}$. Зазвичай часові відліки обираються рівномірно у часі з інтервалом (кроком) дискретизації, наприклад, $T = t_i - t_{i-1}$, де i – довільний номер.

Для отримання дискретного представлення $\{\dots, x_{-1}, x_0, x_1, x_2, \dots\}$ з аналогового сигналу $x(t)$ потрібно знайти скалярний добуток цього сигналу з узагальненою функцією $\eta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$, яка називається дискретизувальною послідовністю.

В результаті отримаємо

$$x_d(t) = \int_{-\infty}^{\infty} x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) dt = \{\dots, x(-T), x(0), x(T), x(2T), \dots\}. \quad (2.12)$$

Отже для дискретизації сигналу виконується операція стробування, тобто обчислення добутку аналогового сигналу $x(t)$ та «гребінчастої» функції $\eta(t)$.

На практиці функція $\eta(t)$ складається із стробувальних імпульсів, які мають кінцеву тривалість. Тому послідовність часових відліків фактично представляє собою модульовану послідовність коротких імпульсів. Імпульсний модулятор має два входи (рис. 2.7).

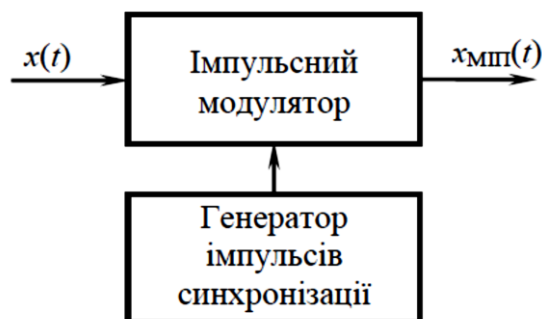


Рис. 2.7. Принцип отримання модульованої імпульсної послідовності

На перший вхід подається вихідний аналоговий сигнал $x(t)$. При потраплянні на його другий вхід послідовності імпульсів синхронізації з періодом T виконується вимірювання миттєвих значень сигналу $x(t)$. В результаті на виході модулятора спостерігається *модульована імпульсна послідовність* $x_{мп}(t)$.

Докладніше особливості модуляції сигналів розглянемо пізніше.

Розглянемо послідовність N дискретних відліків $x(kt)$, заданих на

відрізку $[0, T_1]$, $k = 0, N-1$. Математичною моделлю послідовності таких відліків є амплітудно модульований сигнал зі зсунутих за часом δ -функцій

$$x_{\text{МП}}(t) = T \sum_{k=0}^{N-1} x(kT) \delta(t - kT), \quad (2.13)$$

яка є динамічним представленням сигналу

$$x(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau.$$

Продовжимо подумки цей сигнал з періодом $T_1 = NT$. Оскільки дискретний сигнал (2.13) є періодичним, його спектр має бути дискретним з відстанню між гармоніками $2\pi/(NT)$.

Цей дискретний періодичний сигнал можна представити рядом Фур'є

$$x_{\text{Д}}(t) = \sum_{n=-\infty}^{\infty} \dot{c}(n) e^{jn\omega_1 t}. \quad (2.14)$$

Коефіцієнти $\dot{c}(n)$ ряду (2.14) обчислюються за формулою

$$\begin{aligned} \dot{c}(n) &= \frac{1}{NT} \int_0^{NT} x_{\text{МП}}(t) \exp(-jn\omega_1 t) dt = \frac{1}{NT} \int_0^{NT} \sum_{k=0}^{N-1} x(kT) \delta(t - kT) \exp(-jn\omega_1 t) dt \\ &= \frac{1}{NT} \sum_{k=0}^{N-1} x(kT) \int_0^{NT} \delta(t - kT) \exp(-jn\omega_1 t) dt = \\ &= \frac{1}{N} \sum_{k=0}^{N-1} x(k) \exp(-jn\omega_1 kT). \end{aligned} \quad (2.15)$$

При перетвореннях в (2.15) використана фільтрувальна властивість δ -функції та позначено $x(kT) = x(k)$. Оскільки $\omega_1 = 2\pi/T_1 = 2\pi/NT$, остаточно маємо

$$\dot{c}(n) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi}{N} kn\right). \quad (2.16)$$

Співвідношення (2.16) дозволяє обчислити комплексні амплітуди гармонік дискретного сигналу. Воно є лінійною комбінацією відліків цього сигналу і називається *дискретним перетворенням Фур'є* (ДПФ).

Існує також обернене дискретне перетворення Фур'є

$$x(k) = \frac{1}{N} \sum_{n=0}^{N-1} \dot{c}(n) \exp\left(j \frac{2\pi}{N} kn\right), \quad k = \overline{0, N-1}. \quad (2.17)$$

Зауважимо, що стосовно розміщення множника $1/N$ саме у виразі (5.5) немає повної єдності. У багатьох джерелах цей множник відносять до формули оберненого ДПФ, видаляючи його з формули для прямого ДПФ. В подальшому також буде застосовуватись таке представлення.

Ортогональний дискретний базис Фур'є, в якому виконується ДПФ, є системою дискретних експоненціальних функцій, заданою на дискретній

часовій осі N відліками

$$e_N(k, N) = \exp\left(j \frac{2\pi}{N} kn\right), k = \overline{0, N-1}. \quad (2.18)$$

Система функцій (2.18) є набором N експонент з частотами кратними основній частоті $2\pi/N$, оскільки $e_N(k, N)$ є періодичною по k з періодом N .

Розглянемо основні властивості ДПФ:

1. Дискретне перетворення Фур'є є лінійним перетворенням, тобто якщо послідовностям $x(k)$ і $y(k)$ з одним і тим же періодом N відповідають набори гармонік $\dot{c}_1(n)$ і $\dot{c}_2(n)$, то послідовності $\{ax(k) + by(k)\}$ відповідатиме спектр $\{a\dot{c}_1(n) + b\dot{c}_2(n)\}$. Ця властивість є наслідком лінійності операції підсумовування, яка використовується при обчисленні ДПФ.

2. *Властивість симетрії*, яку має спектр неперервного сигналу, зберігається і для спектру дискретного періодичного сигналу. Якщо відліки $x(k)$ – дійсні числа, тоді коефіцієнти ДПФ, номери яких розташовані симетрично відносно $N/2$, утворюють спряжені пари

$$\begin{aligned} \dot{c}(N-n) &= \frac{1}{N} \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi}{N} k(N-n)\right) = \\ &= \frac{1}{N} \sum_{k=0}^{N-1} x(k) \exp\left(j \frac{2\pi}{N} kn\right) = \dot{c}^*(n) \end{aligned} \quad (2.19)$$

З формули (2.19) виходить, що спектр є спряжено симетричним відносно $N/2$, тобто містить рівно таку ж кількість інформації, що і сам сигнал. Дійсно, якщо початковий сигнал представляється набором з N дійсних чисел, то його спектр представляється набором з $N/2$ комплексних чисел, кожне з яких з інформаційної точки зору еквівалентне двом дійсним. Друга половина спектру взаємно-однозначно пов'язана з першою. Можна вважати, що коефіцієнти $\dot{c}(\frac{N}{2} + 1), \dot{c}(\frac{N}{2} + 2), \dots, \dot{c}(N-1)$ відповідають від'ємним частотам. При вивченні спектру дійсного сигналу вони не дають нової інформації.

Гармоніка з нульовим номером (постійна складова), як випливає з (2.16), є середнім значенням усіх відліків сигналу на одному періоді

$$\dot{c}(0) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) = \dot{c}(N). \quad (2.20)$$

Якщо N – парне число, то

$$\dot{c}(N/2) = \frac{1}{N} \sum_{k=0}^{N-1} x(k)(-1)^k. \quad (2.21)$$

і амплітуда гармоніки з номером $N/2$ визначається сумою відліків зі знаками, що чергуються:

$$\dot{c}(N/2) = \frac{1}{N} [x(0) - x(1) + \dots + x(N-2) - x(N-1)]$$

3. ДПФ кругової згортки

Візьмемо дві послідовності $\{x_1(k)\}$ і $\{x_2(k)\}$ однакової довжини N , ДПФ яких відповідно рівні $\dot{c}_1(n)$ і $\dot{c}_2(n)$. Обчислимо їх кругову згортку на одному періоді

$$y(n) = \frac{1}{N} \sum_{m=0}^{N-1} x_1(m)x_2(n-m). \quad (2.22)$$

Термін «кругова згортка» тут введений у зв'язку з прийнятою періодичністю послідовностей $\{x_1(k)\}$ і $\{x_2(k)\}$. Завдяки цьому значення $x_2(n-m)$, якщо $(n-m) < 0$, дорівнюють $x_2(N+n-m)$, тобто фактично перший і останній члени послідовності йдуть послідовно («коло замикається»).

Знайдемо N -точкове ДПФ цієї згортки з врахуванням властивість зсуву періодичної послідовності

$$\begin{aligned} s(k) &= \frac{1}{N} \sum_{n=0}^{N-1} y(n) \exp\left(-j \frac{2\pi nk}{N}\right) = \frac{1}{N} \sum_{n=0}^{N-1} \left\{ \sum_{m=0}^{N-1} x_1(m)x_2(n-m) \right\} \exp\left(-j \frac{2\pi nk}{N}\right) \\ &= \frac{1}{N} \sum_{m=0}^{N-1} x_1(m) \left\{ \sum_{n=0}^{N-1} x_2(n-m) \exp\left(-j \frac{2\pi nk(n-m)}{N}\right) \right\} \exp\left(-j \frac{2\pi mk}{N}\right) = \dot{c}_1(k)\dot{c}_2(k). \end{aligned} \quad (2.23)$$

Таким чином, круговій згортці дискретизованих сигналів, заданих на одному часовому проміжку, відповідає перемножування їх спектрів.

Обчислення кругової згортки двох сигналів за допомогою ДПФ здійснюється за наступним алгоритмом:

- обчислення ДПФ вихідних сигналів за формулою (2.16);
- перемножування коефіцієнтів отриманих ДПФ згідно (2.23);
- обчислення сигналу $y(n)$ за допомогою оберненого ДПФ (2.17) отриманої послідовності $\{s(k)\}$.

4. Рівність Парсеваля для дискретних сигналів

Визначимо значення $\sum_{n=0}^{N-1} |\dot{c}(n)|^2$, використовуючи формулу ДПФ

$$\sum_{n=0}^{N-1} |\dot{c}(n)|^2 = \sum_{k=0}^{N-1} \left\{ \frac{1}{N} \sum_{k=0}^{N-1} x(k) e_N^*(k, n) \frac{1}{N} \sum_{m=0}^{N-1} x^*(m) e_N(m, n) \right\} = \frac{1}{N} \sum_{k=0}^{N-1} |x(k)|^2. \quad (2.24)$$

Тобто, середня потужність сигналу $x(k)$ на N відліках дорівнює сумі потужностей його частотних компонентів.

5. Зв'язок ДПФ і спектру дискретного сигналу

Маючи один і той самий набір значень дискретного сигналу $x(k)$, можна розрахувати або *спектральну функцію* $\dot{X}(j\omega)$ цього дискретного сигналу по формулі $\dot{X}(j\omega) = \sum_{n=-\infty}^{\infty} x(n) e^{-jn\omega}$, або його ДПФ за формулою (2.16).

Порівняння цих формул показує, що ДПФ є просто дискретними відліками спектральної функції дискретного сигналу, які відповідають частотам $\omega_n = n\omega_d/N$:

$$\dot{c}(n) = X\left(\frac{2\pi n}{NT}\right) = X\left(\omega_d \frac{n}{N}\right). \quad (2.25)$$

Із співвідношення (2.25) маємо важливий висновок: якщо додати до кінцевого набору відліків деяку кількість нулів, спектральна функція дискретного сигналу, природно, не зміниться, але ДПФ дасть більше число спектральних відліків, які відповідатимуть частотам більш щільно розташованим в інтервалі від нуля до частоти дискретизації.

6. Обчислення оберненого ДПФ

Для обчислення оберненого ДПФ можна без будь-яких змін застосовувати пряме ДПФ. Обернене ДПФ N -точкової послідовності $\{\dot{c}(n)\}$, визначається наступним чином

$$x(k) = \sum_{n=0}^{N-1} \dot{c}(n) \exp\left(j \frac{2\pi}{N} kn\right), \quad k = 0, 1, \dots, N-1$$

Якщо взяти комплексне спряження від цього виразу та поділити його на N матимемо

$$\frac{1}{N} x^*(k) = \frac{1}{N} \sum_{n=0}^{N-1} \dot{c}^*(n) \exp\left(-j \frac{2\pi}{N} kn\right). \quad (2.26)$$

Права частина цього виразу є ДПФ послідовності $\{\dot{c}^*(n)\}$, яке може бути обчислене з використанням одного з алгоритмів обчислення прямого ДПФ. Шукану послідовність $\{x(k)\}$ можна отримати, взявши комплексно спряжений з (2.26) вираз і помножити його на N , а саме

$$x(k) = N \left[\frac{1}{N} \sum_{n=0}^{N-1} \dot{c}^*(n) \exp \left(-j \frac{2\pi}{N} kn \right) \right]^* \quad (2.27)$$

В подальшому будемо користуватися більш поширеним представленням прямого та оберненого перетворення Фур'є, коли множник $1/N$ застосовується у співвідношенні для обчислення оберненого перетворення Фур'є.

2.5. Швидке перетворення Фур'є

Обчислення перетворення Фур'є, тобто знаходження спектральних складових $\dot{c}(n)$ дискретного комплексного сигналу $x(k)$ безпосередньо по формулі ДПФ вимагає N^2 комплексних множень і $N(N-1)$ комплексних складань. Оскільки кількість обчислень, а отже, і час обчислень приблизно є пропорційними N^2 , то при великих N тривалість операції є дуже великою. Тому знаходження спектру в реальному часі вказаними методами навіть для сучасної обчислювальної техніки є складним завданням.

Але існують обчислювальні процедури, які зменшують кількість множень і складань. Основний принцип усіх цих алгоритмів полягає в розкладанні операцій обчислення ДПФ сигналу довжини N на обчислення перетворень Фур'є з меншим числом точок. Розділивши аналізований набір відліків на частини, обчислюють їх ДПФ і об'єднують результати. Такі процедури дістали назву алгоритмів швидкого перетворення Фур'є (ШПФ, *Fast Fourier Transform, FFT*).

При реалізації ШПФ можливі декілька варіантів організації обчислень залежно від способу ділення послідовності відліків на частини (*проріджування за часом (decimation in time, DIT)* або *по частоті (decimation in frequency, DIF)*) і від того, на скільки фрагментів розбивається послідовність на кожному кроці (*основа ШПФ*). Найбільш простими і широко використовуваними є алгоритми ШПФ з основою 2, коли довжина послідовності $x(k)$ є цілою мірою числа 2, тобто $N=2^v$, де v – ціле число.

Проаналізуємо більш детально операції ДПФ на прикладі $N=8$. Введемо позначення для дискретної експоненціальної функції [7]

$$\exp \left(-j \frac{2\pi}{N} kn \right) = w_N^{nk}. \quad (2.28)$$

Для $N=8$ маємо $w = \exp \left(-j \frac{2\pi}{8} 1 \cdot 1 \right) = \exp \left(-j \frac{\pi}{4} \right)$ (тут і далі індекс $N=8$

опущено для простоти запису).

Вираз для обчислення ДПФ (вже без множника $1/N$)

$$\begin{aligned} \hat{c}(n) = c_n = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi}{N} kn\right) = \sum_{k=0}^{N-1} x(k) w^{kn}, \\ n, k \in 0, 1, \dots, N-1 \end{aligned} \quad (2.29)$$

для більшої наочності представимо у вигляді добутку двох матриць

$$\mathbf{C} = \mathbf{W}\mathbf{X}, \quad (2.30)$$

де квадратна матриця ДПФ має вигляд (тут елементи матриці формуються по рядках (n) і стовпцях (k) при зміні n, k від 0 до 7; 1-й рядок і стовпець дорівнюють 1, оскільки n або $k=0$; індекс в позначенні елемента матриці утворюється добутком $n \cdot k$)

$$\mathbf{W} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 & w^4 & w^5 & w^6 & w^7 \\ 1 & w^2 & w^4 & w^6 & w^8 & w^{10} & w^{12} & w^{14} \\ 1 & w^3 & w^6 & w^9 & w^{12} & w^{15} & w^{18} & w^{21} \\ 1 & w^4 & w^8 & w^{12} & w^{16} & w^{20} & w^{24} & w^{28} \\ 1 & w^5 & w^{10} & w^{15} & w^{20} & w^{25} & w^{30} & w^{35} \\ 1 & w^6 & w^{12} & w^{18} & w^{24} & w^{30} & w^{36} & w^{42} \\ 1 & w^7 & w^{14} & w^{21} & w^{28} & w^{35} & w^{42} & w^{49} \end{pmatrix}. \quad (2.31)$$

а сигнал має вигляд однорядкової матриці (вектору) $\mathbf{X} = (x_0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7)$.

Проаналізуємо w^n . Точки $w^n = e^{-j\frac{\pi}{4}n}$ знаходяться на колі одиничного радіусу комплексної площини і є періодичною послідовністю (через $\pi/4$). Тобто значення $n = \overline{0, N-1}$ повністю визначають будь-яку степінь при $n \geq 8$ (за умови $N=8$) (рис.2.8).

На рис. 2.8 показані числа

$$w^n = e^{-j\frac{\pi}{4}n} = \cos\left(\frac{\pi}{4}n\right) - j \cdot \sin\left(\frac{\pi}{4}n\right).$$

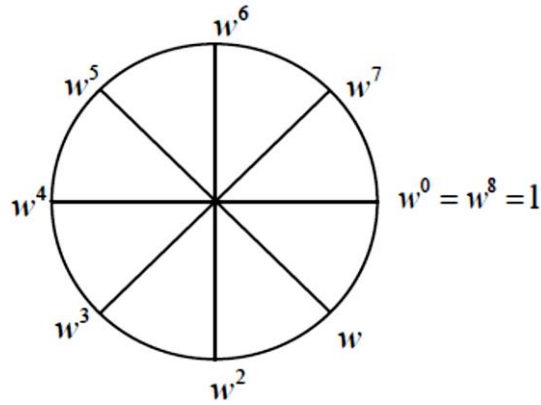


Рис. 2.8. Комплексна площина w

Звернемо також увагу на те, що числа, які відповідають точкам розташованим діаметрально протилежно на колі відрізняються лише знаком.

Тому $w^4 = -w^0 = -1, w^5 = -w, w^6 = -w^2, w^6 = -w^2, w^7 = -w^3$. Тобто будь-яка ціла степінь w може бути представлена через $1, w, w^2, w^3$. З урахуванням цього матрицю ДПФ можна подати у вигляді

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & w & w^2 & w^3 & -1 & w & -w^2 & -w^3 \\ 1 & w^2 & -1 & -w^2 & 1 & w^2 & -1 & -w^2 \\ 1 & w^3 & -w^2 & w & -1 & -w^3 & w^2 & -w \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -w & w^2 & -w^3 & -1 & w & -w^2 & w^3 \\ 1 & -w^2 & -1 & w^2 & 1 & -w^2 & -1 & w^2 \\ 1 & -w^3 & -w^2 & -w & -1 & w^3 & w^2 & w \end{pmatrix}. \quad (2.32)$$

В кожному рядку цієї матриці є або однакові або такі, що різняться за знаком члени. Ідея ШПФ застосовує описані вище властивості кругової згортки і, крім того, передбачає скорочення кількості операцій множення відповідно до тотожності

$$a \cdot b \pm a \cdot c = a \cdot (b \pm c). \quad (2.33)$$

Ліворуч у виразі (5.22) два множення, а праворуч лише одне. З урахуванням цього знайдемо складові дискретного спектру (перемноження матриць за правилом «рядок (вектор \mathbf{X}) на стовпець (матриця \mathbf{W}):

$$c_0 = [(x_0 + x_4) + (x_2 + x_6)] + [(x_1 + x_5) + (x_3 + x_7)];$$

$$c_1 = [(x_0 - x_4) + (x_2 - x_6) \cdot w^2] + [(x_1 - x_5) + (x_3 - x_7) \cdot w^2] \cdot w;$$

$$\begin{aligned}
c_2 &= [(x_0 + x_4) - (x_2 + x_6)] + [(x_1 + x_5) - (x_3 + x_7)] \cdot w^2; \\
c_3 &= [(x_0 - x_4) + (x_2 - x_6) \cdot w^2] + [(x_1 - x_5) - (x_3 - x_7) \cdot w^2] \cdot w^3; \\
c_4 &= [(x_0 + x_4) + (x_2 + x_6)] - [(x_1 + x_5) + (x_3 + x_7)]; \\
c_5 &= [(x_0 - x_4) + (x_2 - x_6) \cdot w^2] - [(x_1 - x_5) + (x_3 - x_7) \cdot w^2] \cdot w; \\
c_6 &= [(x_0 + x_4) - (x_2 + x_6)] - [(x_1 + x_5) - (x_3 + x_7)] \cdot w^2; \\
c_7 &= [(x_0 - x_4) + (x_2 - x_6) \cdot w^2] - [(x_1 - x_5) - (x_3 - x_7) \cdot w^2] \cdot w^3.
\end{aligned}$$

Видно, що у круглих дужках присутні лінійні *комбінації* елементів вектора \mathbf{X} . Кількість таких комбінацій дорівнює 8. В квадратних дужках вагові *комбінації сум* у круглих дужках (їх теж вісім). Останні операції передбачають утворення вагових сум з результатів у квадратних дужках.

Ці три послідовні етапи обчислення можуть бути представлені у вигляді добутку трьох порівняно простих матриць (відповідно з права на ліво). Таким чином, матриця ДПФ набуде вигляду:

$$\mathbf{W} = \mathbf{W}_3 \mathbf{W}_2 \mathbf{W}_1, \quad (2.34)$$

$$\mathbf{W}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}. \quad (2.35)$$

$$\mathbf{W}_2 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & w^2 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & w^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & w^2 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & -w^2 \end{pmatrix}. \quad (2.36)$$

$$W_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & w & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & w^2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & w^3 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -w & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & w^2 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -w^3 \end{pmatrix}. \quad (2.37)$$

Якщо обчислити добуток цих трьох матриць отримаємо матрицю W . Розкладання матриць на прості множники виду (2.34) називають *факторизацією*. Процедуру обчислення дискретного спектру (коефіцієнтів c_n) зручно подати у вигляді графа (рис. 2.9), який пояснює послідовність дій, виконуваних факторизованою матрицею.

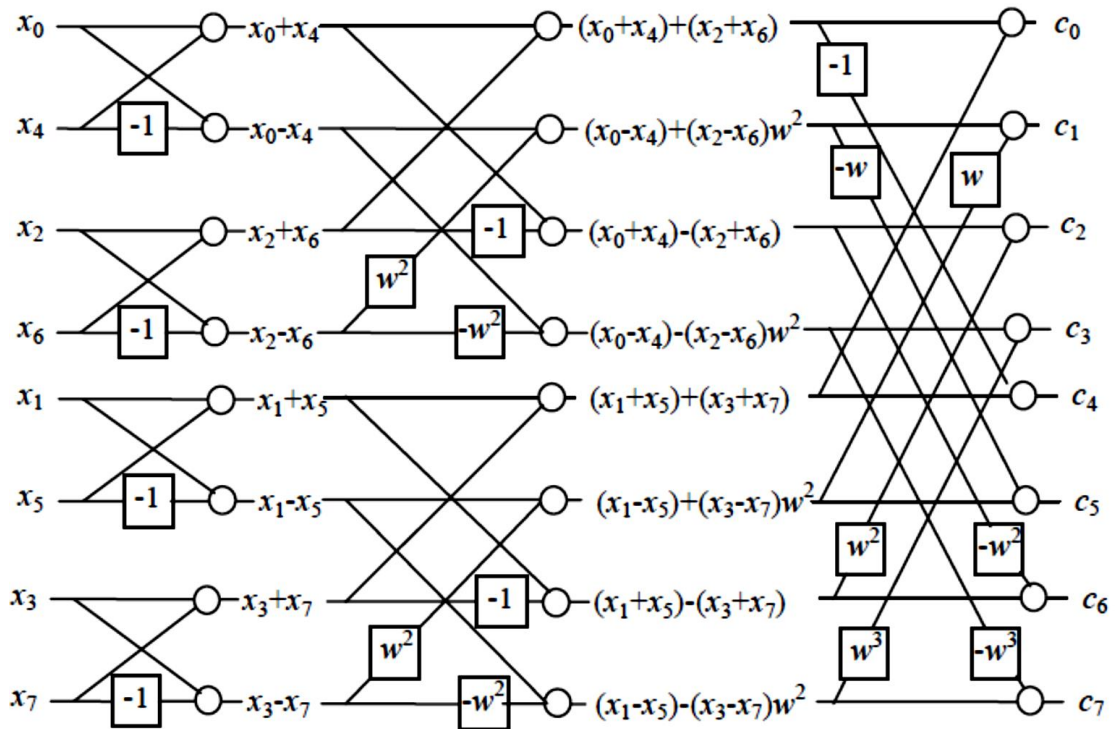


Рис. 2.9. Граф ШПФ. Кругом позначено складання, а квадратом – множення на величину, що вказана в квадраті

Промисловість серійно випускає інтегральні мікросхеми, які виконують чотирьохточкове ДПФ, яке є базовою операцією. Цю мікросхему називають «метеликом». Вона дозволяє виготовляти спеціалізовані процесори ШПФ для сигналів з довільним N . Алгоритмічна схема

«метелика» і її позначення наведені на рис. 2.10.

Кожний блок, що реалізує об'єднання двох ДПФ, має два вхідних і два вихідних сигнали. Один із вхідних сигналів множиться на комплексну експоненту w^n , після чого додається до другого вхідного сигналу або віднімається від нього, формуючи таким чином два вихідні сигнали.

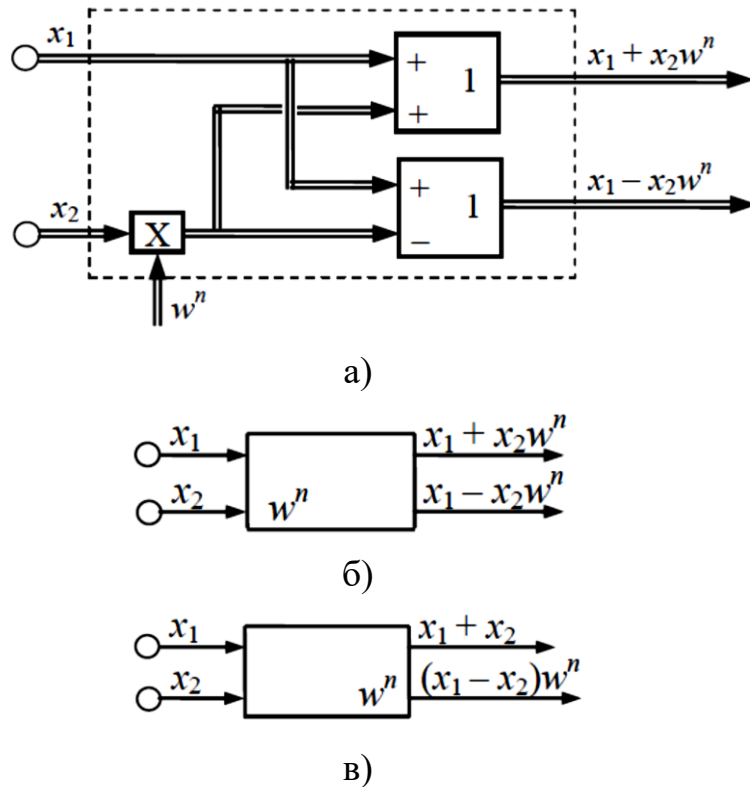


Рис. 2.10. а) алгоритмічна схема ШПФ з проріджуванням за часом, б) умовне позначення «метелика» ШПФ з проріджуванням за часом, в) умовне позначення «метелика» ШПФ з проріджуванням за частотою

Якщо на входи мікросхеми надходять сигнали комплексних чисел x_1 і x_2 , то на виході формуються числа згідно з виразами $x_1 + x_2 w^n$ і $x_1 - x_2 w^n$, які є базовою операцією алгоритму з проріджуванням за часом. Для алгоритму з проріджуванням за частотою базова операція має вигляд: $x_1 + x_2$ і $(x_1 - x_2) w^n$.

Оскільки комплексна величина є парою чисел (реальна і уявна частини), то практично «метелик» має чотири входи і чотири виходи.

Як видно зі структурної схеми восьмиточкового спеціалізованого процесора ШПФ реалізованого на «метеликах» (рис. 2.11), в ній реалізовані операції визначені факторизованою матрицею (2.34).

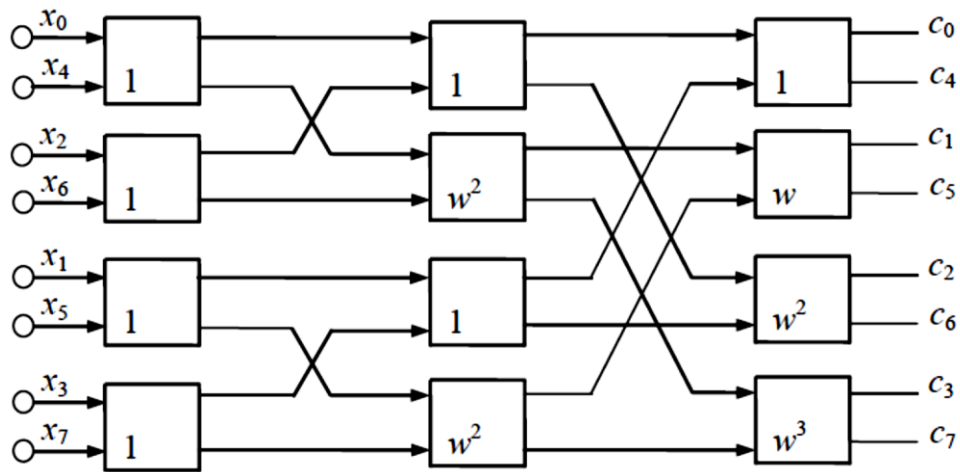


Рис. 2.11. Структурна схема спецпроцесора ШПФ на «метеликах»

Формально ШПФ по основі 2 здійснюється шляхом розбивання вихідної послідовності на парні (x_0, x_2, x_4, \dots) і непарні (x_1, x_3, x_5, \dots) підпослідовності. З ними далі поступають таким же чином. Ці процедури повторюються доти, поки не залишиться по одній парі чисел, які поступають на «метелики» першого етапу. Для восьмиточкового ШПФ цими парами є $x_0x_4, x_2x_6, x_1x_5, x_3x_7$. На першому етапі операцій множення практично не проводиться (множення на ± 1). Тут формуються суми і різниці отриманих пар.

Далі алгоритм обчислення відповідає рис. 2.11. Він складається з трьох етапів. В загальному ж випадку для N -точкового ШПФ потрібно $\log_2 N$ етапів. На кожному з них виконується $N/2$ операцій множення. Таким чином, алгоритм ШПФ потребує $\frac{N}{2} \log_2 N$ операцій множення.

Загальний алгоритм ДПФ має N^2 операцій множення. Тому алгоритм ШПФ порівняно з алгоритмом ДПФ дає вигреш по кількості операцій множення в $\frac{2N}{\log_2 N}$ разів. В цілому, загальна кількість операцій при обчисленні ШПФ складає приблизно $\frac{N}{2} \log_2 N$ множень і $N \log_2 N$ складань комплексних чисел.

Таким чином:

- ШПФ не є наближеним алгоритмом, при відсутності похибок у розрахунках він дасть точно такий же результат, що і пряма формула ДПФ. Прискорення досягається виключно за рахунок оптимальної організації

розрахунків;

- застосування ШПФ має сенс, якщо число елементів в послідовності, що аналізується є степенем числа 2;

- алгоритм ШПФ призначений для одночасного розрахунку всіх спектральних відліків $\hat{c}(n)$. Якщо ж необхідно отримати ці відліки лише для деяких n , може виявитись більш прийнятним застосування прямого ДПФ.

Часто ДПФ використовують для дослідження спектра сигналу. Зазвичай при цьому найкориснішою є інформація про амплітудний, а не фазовий спектр.

(Часто шкалу амплітуд градуують в децибелах, які вказують не на величину певної амплітуди, а на відношення двох величин – виміряної A_n та базової A_0 : $A_{dB} = 10 \lg \frac{A_n}{A_0}$. Відповідно, відношення амплітуд 10^x відповідає значенню 10 дБ, 100^x – 20 дБ, 1000^x – 30 дБ тощо).

Перед обчисленням спектру вибирається певний відрізок сигналу, де потрібно обчислити спектр. Як правило, довжина відрізка має бути степенем двійки. Інакше сигнал потрібно доповнити нулями до потрібної довжини. Далі обчислюється ДПФ для цього відрізка.

Але при розкладанні функції в ряд Фур'є вважається, що функція є періодичною з періодом, що дорівнює розміру ШПФ. Тобто обчислюється спектр періодичної функції, а не вибраного відрізка. При цьому на границях періодів така функція має розриви, які спотворюють спектр. Для усунення спотворення використовуються вагові вікна, які плавно зменшують значення функції на кінцях вибраного відрізка [5, 6].

Вагові вікна мають форму, близьку до гаусоїди і при перемноженні на відліки функції усувають розриви. Існують багато вагових функцій. Вигляд двох широко вживаних вікон показано на рис.2.12

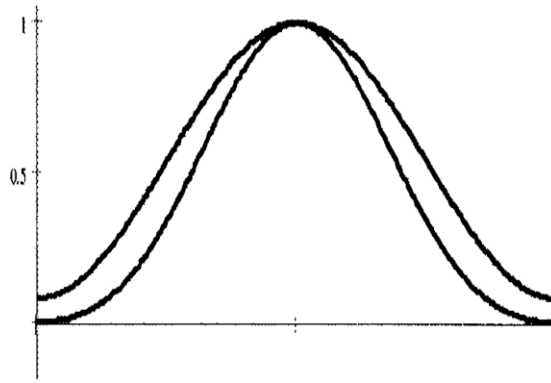


Рис. 2.12. Вагові вікна Хеммінга (верхнє) та Блекмана (нижнє)

2.6. Пряме та обернене z -перетворення, його властивості

Зручним способом аналізу дискретних послідовностей є z -перетворення. При z -перетворенні різницеві рівняння, що описують роботу дискретної системи, перетворюються в рівняння алгебраїчні, з якими простіше виконувати необхідні дії. Z -перетворення відіграє для дискретних сигналів і систем таку ж роль, як перетворення Лапласа для аналогових сигналів.

Перетворення Лапласа здебільшого використовується для рішення диференціальних рівнянь шляхом перетворення їх на алгебраїчні рівняння. Пряме перетворення Лапласа ставить у відповідність функції $f(t)$ дійсної змінної t (сама функція може бути і комплексною) комплексну функцію $F(s)$ комплексної змінної $s = \sigma + j\omega$ (σ – дійсне число, ω – частота в радіанах за секунду):

$$F(s) = \int_0^{\infty} f(t)e^{-st} dt . \quad (2.38)$$

Функція комплексного аргументу $F(s)$ називається *зображенням Лапласа* (або інтегральним перетворенням Лапласа) для функції-оригіналу $f(t)$. Оскільки функція e^{-st} є безрозмірною, то змінна s має розмірність частоти (обернено до часу) і іноді називається комплексною частотою.

Суть формули (2.38) полягає в тому, що функція $f(t)$ поточково перемножається на функцію e^{-st} при різних значеннях s . Після цього знаходять площу під графіком функції $f(t)e^{-st}$ шляхом додавання всіх добутків. Залежність цієї площі від кожного значення $s = \sigma + j\omega$ буде значенням перетворення Лапласа $F(s)$.

Однозначність та можливість існування прямого перетворення Лапласа

визначається наступними достатніми умовами:

1) Умова однозначності інтегралу: для $t < 0$ всі можливі функції $f(t)=0$. При невиконанні цієї умови функції, що відрізняються між собою на від'ємній дійсній напівосі, матимуть однакові Лаплас-образи.

2) Умови існування інтегралу:

а) для $t \geq 0$ на довільному скінченному проміжку дійсної осі функція $f(t)$ має не більше ніж скінчену кількість розривів першого роду;

б) при $t \rightarrow \infty$ функція $|f(t)| \leq A \exp(at)$, тобто має скінчений показник степеню зростання $a > 0$ ($A > 0$ – довільна стала). Це означає, що функція може зростати, при наближенні до нескінченно віддаленої точки, але це зростання не швидше за експоненціальне.

Більш загальне (двостороннє) перетворення Лапласа має нижню границю $-\infty$, але для т.зв. *каузальних* систем (які реалізуються фізично) немає значення, який стан мала система до моменту $t = 0$, тому використовуватимемо одностороннє перетворення.

Визначення z-перетворення. Дискретній послідовності відліків $\{x(n)\}$ ставиться у відповідність функція комплексної змінної z , яка визначається таким чином

$$Z[x(n)] = X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}. \quad (2.39)$$

Функція $X(z)$ визначена тільки для тих значень z , при яких ряд (2.39) сходиться. Якщо послідовність $x(n)$ має обмежену довжину, то $X(z)$ сходиться в z -площині скрізь, за виключенням, можливо, точок $z = 0$ або $z = \infty$.

Отримаємо z -перетворення для деяких дискретних сигналів, що часто зустрічаються на практиці.

Одиничний імпульс, який визначається як

$$u_0(n) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0. \end{cases} \quad (2.40)$$

Використовуючи формулу (2.39), отримаємо:

$$X_0(z) = \sum_{n=-\infty}^{\infty} u_0(n)z^{-n} = 1. \quad (2.41)$$

Функція $X_0(z)$ сходиться в усій комплексній площині.

Одиничний стрибок, який визначається як

$$u_1(n) = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0. \end{cases} \quad (2.42)$$

Використовуюючи визначення z -перетворення, отримаємо

$$X_1(z) = \sum_{n=-\infty}^{\infty} u_1(n)z^{-n} = \sum_{n=0}^{\infty} z^{-n}. \quad (2.43)$$

Ряд (2.43) є сумою нескінченної геометричної прогресії з першим членом 1 і знаменником z^{-1} . Як відомо, такий ряд сходиться при $|z^{-1}| < 1$, тобто при $|z| > 1$ і його сума складає

$$X_1(z) = \frac{1}{1-z^{-1}} = \frac{z}{z-1}. \quad (2.44)$$

Значення $z=1$ є єдиною особливою точкою (полюсом) функції $X_1(z)$.

Експоненціальна дискретна функція

$$u_2(n) = \begin{cases} a^n, & n \geq 0, \\ 0, & n < 0. \end{cases} \quad (2.45)$$

Обчислимо її z -перетворення

$$X_2(z) = \sum_{n=-\infty}^{\infty} u_2(n)z^{-n} = \sum_{n=0}^{\infty} a^n z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n. \quad (2.46)$$

Як і у попередньому випадку ряд (2.46) є сумою геометричної прогресії з першим членом, рівним 1 і знаменником az^{-1} . Таким чином, ряд сходиться при $|az^{-1}| < 1$ тобто при $z > a$, і має особливу точку при $z = a$. Сума ряду становить

$$X_2(z) = \frac{1}{1-az^{-1}} = \frac{z}{z-a}. \quad (2.47)$$

Зв'язок z -перетворення з перетворенням Лапласа можна знайти, записавши аналоговий сигнал у вигляді суми добутків дискретних відліків і набору дельта-функцій

$$x(t) = \sum_{n=-\infty}^{\infty} x(n)\delta(t - nT), \quad (2.48)$$

де T – крок дискретизації.

Перетворення Лапласа для такого сигналу дорівнює

$$F(s) = \int_0^{\infty} x(t)e^{-st} dt = \int_0^{\infty} \sum_{n=-\infty}^{\infty} x(n)\delta(t - nT)e^{-st} dt = \sum_{n=-\infty}^{\infty} x(n) \int_0^{\infty} \delta(t - nT)e^{-st} dt. \quad (2.49)$$

Скориставшись фільтрувальною властивістю дельта-функції, маємо

$$F(s) = \sum_{n=-\infty}^{\infty} x(n)e^{-snT} \quad (2.50)$$

З порівняння співвідношень (2.39) і (2.50), видно, що одна формула переходить в іншу при заміні $z = e^{sT}$. Таким чином, z -перетворення можна отримати з перетворення Лапласа шляхом переходу до нової змінної:

$$F(p) \Rightarrow X(z) \quad \text{при } z = e^{sT}. \quad (2.51)$$

Перехід від перетворення Лапласа до z -перетворення при описанні дискретних систем потрібний з наступної причини.

Дискретизація аналогового сигналу призводить до періодичності частотного спектру, тобто появи нескінченного ряду зсунутих копій спектру початкового неперервного сигналу. Очевидно, ефект дискретизації веде до появи в площині $s = \sigma + j\omega$ нескінченної конфігурації особливих точок (поліосів і нулів).

При переході від s -площини до z -площини точка $p_k = \sigma_k + j\omega_k$ відображується в точку $z_k = e^{(\sigma_k + j\omega_k)T} = e^{\sigma_k T} e^{j\omega_k T}$.

Тому шлях уздовж уявної осі $j\omega$ в s -площині відображується в одиничне коло в z -площині, оскільки на уявній осі $\sigma_k = 0$ і, отже, $z = 1$. При цьому ліва (стійка) смуга s -площини шириною $\Omega = 2\pi/T$ відображується всередину кола одиничного радіусу z -площини. Усі наступні стійкі смуги s -площини шириною Ω , які відповідають періодичному частотному спектру дискретного сигналу, також відображуються всередину круга одиничного радіусу z -площини. Тому конфігурація особливих точок (поліосів і нулів) в z -площині стає кінцевою.

Таким чином, взаємна відповідність між z -перетворенням $X(z)$ та перетворенням Лапласа $F(s)$ описується наступним чином:

$$X(z) = F\left(\frac{1}{T} \ln z\right); \quad F(s) = X(e^{sT}). \quad (2.52)$$

Подібними формулами описується зв'язок z -перетворення $X(z)$ з перетворенням Фур'є $F(\omega)$ (при цьому немає необхідності вважати послідовність однобічною):

$$X(z) = F\left(\frac{1}{jT} \ln z\right); \quad F(j\omega) = X(e^{j\omega T}). \quad (2.53)$$

2.7. Зв'язок ДПФ з z -перетворенням

Порівнюючи формулу прямого ДПФ дискретної послідовності $x(n)$ з формулою z -перетворення, побачимо, що коефіцієнти ДПФ $\hat{x}(k)$ дорівнюють значенням z -перетворення цього сигналу в N точках, рівномірно розподілених по одиничному колу z -площини. Ці коефіцієнти однозначно представляють саму послідовність, оскільки вона може бути

точно відновлена за допомогою оберненого ДПФ.

Отримаємо z -перетворення послідовності через коефіцієнти ДПФ цієї послідовності

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n} = \frac{1}{N} \sum_{n=0}^{N-1} \left[\sum_{k=0}^{N-1} c(k) e^{j\frac{2\pi nk}{N}} \right] z^{-n} =$$

$$\frac{1}{N} \sum_{k=0}^{N-1} c(k) \sum_{n=0}^{N-1} \left(e^{j\frac{2\pi k}{N}} z^{-1} \right)^n = \frac{1}{N} \sum_{k=0}^{N-1} c(k) \frac{1 - z^{-N}}{1 - e^{j\frac{2\pi k}{N}} z^{-1}}. \quad (2.54)$$

Формула (2.54) показує, що z -перетворення кінцевої послідовності $x(n)$, $n = \overline{0, N-1}$, безпосередньо пов'язане з коефіцієнтами $c(k)$, $k = \overline{0, N-1}$, її ДПФ.

Властивості z -перетворення:

1. *Лінійність.*

Якщо $X_1(z)$ і $X_2(z)$ є z -перетвореннями відповідних сигналів $x_1(n)$ і $x_2(n)$, то сигналу $y(n) = ax_1(n) + bx_2(n)$ відповідатиме z -перетворення $Y(z) = aX_1(z) + bX_2(z)$ при будь-яких постійних a і b .

2. *Затримка (зсув послідовності).*

Якщо z -перетворення сигналу $x(n)$ дорівнює $X(z)$, то z -перетворення сигналу $y(n) = x(n - n_0)$, затриманого на n_0 тактів, дорівнює $X(z) \cdot z^{-n_0}$.

3. *Теорема множення на степеневу функцію*

Якщо z -перетворення сигналу $x(n)$ дорівнює $X(z)$, то z -перетворення сигналу $y(n) = a^n x(n)$ дорівнює $X(a^{-1}z)$.

4. *Теорема множення на n*

Якщо z -перетворення сигналу $x(n)$ дорівнює $X(z)$, то z -перетворення сигналу $y(n) = nx(n)$ дорівнює $-z \frac{dX(z)}{dz}$.

5. *Згортка*

Введемо дискретну лінійну згортку $w(n)$, яку визначимо таким чином:

$$w(n) = \sum_{k=-\infty}^{\infty} x(k)y(n-k) = \sum_{k=-\infty}^{\infty} y(k)x(n-k). \quad (2.55)$$

Обчислимо її z -перетворення

$$\begin{aligned}
W(z) = Z[w(n)] &= Z\left[\sum_{k=-\infty}^{\infty} x(k)y(n-k)\right] = \sum_{n=-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} x(k)y(n-k)z^{-n}\right] = \\
&= \sum_{n=-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} x(k)y(n-k)z^{-k}z^{-(n-k)}\right] = \sum_{k=-\infty}^{\infty} \left[x(k)z^{-k} \sum_{n=-\infty}^{\infty} y(n-k)z^{-(n-k)}\right] =
\end{aligned}
\tag{2.56}$$

Отже, z -перетворення лінійної згортки двох дискретних сигналів дорівнює добутку їх z -перетворень.

Різницеві рівняння зазвичай визначені при $n \geq 0$ і мають набір початкових умов. Тому при рішенні практичних задач зазвичай вводять одностороннє z -перетворення, що має вигляд

$$X(z) = \sum_{n=0}^{\infty} x(n)z^{-n}. \tag{2.57}$$

Для багатьох сигналів властивості одностороннього z -перетворення аналогічні властивостям звичайного z -перетворення. Основним виключенням є властивість, пов'язана із зсувом (затримкою) сигналів. Так, затримка на один відлік, як і раніше, призводить до множення одностороннього z -перетворення на $-1z$, але при цьому необхідно врахувати значення сигналу $n < 0$, тобто початкові умови.

6. Властивість коефіцієнтів розкладання z -перетворення в степеневий ряд

Якщо z -перетворення сигналу $x(n\Delta t)$ дорівнює $X(z) = q_0 + q_1z^{-1} + q_2z^{-2} + q_3z^{-3} + \dots$, то коефіцієнти q_0, q_1, q_2, \dots розкладання $X(z)$ в степеневий ряд по z дорівнюють значенням відліків функції $x(nT)$, тобто $q_n = x(nT)$.

7. Теорема початкового і кінцевого (фінального) значень

Є важливими для з'ясування поведінки дискретної послідовності $x(n)$ відповідно до відомого її z -перетворення. Крім того, теореми початкового і кінцевого значень дозволяють контролювати обчислення оберненого z -перетворення.

Теорема початкового значення встановлює, що

$$x(0) = \lim_{z \rightarrow \infty} X(z)$$

Теорема кінцевого значення встановлює, що

$$x(\infty) = \lim_{z \rightarrow 1} (1 - z^{-1}) X(z).$$

2.8. Обернене z-перетворення

Відповідно до формули одностороннього z-перетворення (2.57) функція $X(z)$ визначає всю нескінченну сукупність відліків $\{x(0), x(1), \dots, x(n), \dots\}$. Помножимо обидві частини ряду (2.57) на множник z^{n-1} :

$$z^{n-1}X(z) = x(0)z^{n-1} + x(1)z^{n-2} + \dots + x(n)z^{-1} + \dots, \quad (2.58)$$

а потім обчислимо інтеграли від обох частин отриманої рівності, взявши за контур інтегрування замкнуту криву, що лежить цілком в області аналітичності і охоплює усі полюси функції $X(z)$. При цьому скористаємось відомими співвідношеннями для інтегралу по замкнутому контуру C , в середині якого лежить точка $z = 0$:

$$\oint_C z^n dz = \begin{cases} 2\pi j, & n = -1, \\ 0, & n \neq -1. \end{cases}$$

Видно, що інтеграли від усіх доданків правої частини (2.58), окрім доданку з номером n , будуть рівними нулю. Тому

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz = Z^{-1}[X(z)], \quad (2.59)$$

де C – замкнутий контур $C > R_1$, R_1 – радіус кола збіжності $X(z)$.

Вираз (2.59) називають оберненим z-перетворенням, воно дозволяє знайти відліки $x(n)$ по z-зображенню $X(z)$. Обернене z-перетворення існує тільки для таких функцій $X(z)$, які мають кінцеве число особливих точок (полюсів), причому особливість в кожній з них є усувною.

2.9. Загальні відомості про вейвлет-перетворення

Перетворення Фур'є та ряди Фур'є є потужним математичним апаратом для фізичної інтерпретації процесів при аналізі сигналів, що їх характеризують. Однак іноді вони виявляються недостатньо ефективними. Так, щоб одержати спектральну інформацію на обраній частоті, необхідно мати і минулу, і майбутню часову інформацію. До того ж перетворення Фур'є не враховує, що частота сигналу може змінюватись згодом. Воно, наприклад, не відрізняє сигнал, що є сумою двох синусоїд з різними частотами, від сигналу, що складається з тих же синусоїд, що йдуть послідовно одна за іншою.

Для одержання високочастотної інформації з високою точністю

важливо «витягати» її з відносно малих часових інтервалів, а не із усього сигналу; і навпаки – низькочастотну спектральну інформацію потрібно витягати з відносно протяжних часових інтервалів сигналу.

Частина описаних труднощів знімається при використанні віконного перетворення Фур'є. Однак нескінченно осцилююча базисна функція (синусоїда) не дозволяє одержувати по-справжньому локалізовану інформацію.

Вейвлет-аналіз став фактично стандартним інструментом для дослідження частотно-часової поведінки сигналів. На відміну від класичних методів Фур'є-аналізу, вейвлет-аналіз дає опосередковану інформацію про частотний склад сигналу та його зміну в часі. Це досягається через розклад сигналу на сукупність складових з компактним носієм, кожний з яких є розтягнутою або стиснутою копією єдиної материнської вейвлет-функції.

Вейвлет-перетворення необхідне для аналізу структури і поведінки сигналів, які містять у собі ділянки різної тривалості з коливаннями різної швидкості або різної частотної наповненості. Часто в літературі такі сигнали називають «нестационарними», маючи на увазі, що їх спектральні характеристики досить істотно змінюються з часом, причому ці зміни можуть відбуватися протягом різного періоду. Для такого випадку, коли в сигналі є локалізовані в часі ділянки різних частотних властивостей, використовується розклад у набір вейвлет-функцій, які самі є також локалізованими в часі і мають різний спектральний склад. Такий розклад гарантує проведення локального аналізу сигналу: якщо певний коефіцієнт розкладу сигналу є великим, то можна визначити ту ділянку часу, якій він відповідає, а отже, проаналізувати її більш детально і визначити, яким є частотний склад тієї ділянки сигналу та її тривалість. Скрізь, де треба дослідити нерегулярності в сигналах, альтернативи вейвлет-перетворенням майже немає.

Внаслідок названих унікальних можливостей різноманітні модифікації вейвлет-аналізу стали дуже поширеними в практиці обробки та аналізу сигналів. Вони використовуються для аналізу біомедичних сигналів та зображень, в геології та радіолокації, для стиснення даних, адаптивної фільтрації, кодування і внесення в сигнали цифрових водяних знаків,

застосовуються в цифровому зв'язку і телебаченні високої чіткості та розпізнаванні образів.

Докладну теорію вейвлет-перетворень викладено в багатьох джерелах; класичними підручниками є [8, 9].

При проведенні аналізу сигналів за допомогою неперервного вейвлет-перетворення (НВП) основним етапом є розрахунок вейвлет-коефіцієнтів. Для розрахунку вейвлет-перетворень сигналів переважно використовується апроксимація неперервного вейвлет-перетворення за допомогою ортогонального *діадного* вейвлет-перетворення або застосовується *кратномасштабний аналіз*. При цьому необхідно мати формульний вираз для розрахунку значень материнського вейвлету, до того ж, материнська функція має утворювати при масштабуванні ортогональний базис.

Але часто для аналізу використовується материнська функція, яка не має точного математичного опису.

На підставі поняття про векторний простір загальноприйнятим підходом до аналізу сигналів $s(t)$ стало їхнє представлення у вигляді зваженої суми простих складових – базисних функцій $\psi_k(t)$, помножених на коефіцієнти C_k :

$$s(t) = \sum_k C_k \psi_k(t).$$

Оскільки базисні функції $\psi_k(t)$ передбачаються заданими як функції цілком певного виду, то тільки коефіцієнти C_k містять інформацію про конкретний сигнал. Таким чином, можна говорити про можливість представлення довільних сигналів на основі рядів з різними базисними функціями.

Наприклад, в ряді Фур'є $s(t) = \sum_k C_k \exp(jtk)$ використовуються в якості базисної функції синусоїди $\exp(jtk) = \cos(tk) + j\sin(tk)$. Вони гранично локалізовані в частотній області й взагалі не локалізовані в часовій області.

Термін «вейвлет» (wavelet) у перекладі з англійської мови означає «коротка або маленька хвиля». У побут його ввів сейсмолог Морле, тому що подібні хвилі поширюються в об'ємі Землі при тектонічних процесах.

Термін «вейвлет» є вдалим через відображення в ньому суті процесу, тому що більшість вейвлетів має часові залежності з яскраво вираженим

коливальним компонентом, як у хвилях. Вейвлети займають проміжне положення між наведеними вище базисними функціями – синусоїдою та функцією Кронекера.

Досить грубо можна уявити вейвлети як деякі хвильові функції, здатні здійснювати перетворення Фур'є не по всій часовій осі, а локально по місцю свого розташування. Для цього, природно, крім зміни «середньої частоти» вейвлети повинні переміщатися до того місця сигналу або функції, у якому повинне здійснюватися «локальне перетворення Фур'є» і масштабуватися.

Така інтерпретація вейвлетів є дуже спрощеною й дає про них тільки саме грубе уявлення, яке часом навіть принципово помилкове через те, що переважна більшість вейвлетів не є модульованою по амплітуді синусоїдальною хвилею, а властивості вейвлетів принципово та сильно відрізняються від властивостей синусоїди.

Базисними функціями вейвлетів можуть бути різні функції, у тому числі такі, що нагадують модульовані синусоїди, функції зі стрибками рівня тощо. Це забезпечує різне представлення сигналів з локальними стрибками та розривами наборами вейвлетів різних типів і дає можливість вибору найбільш прийнятних вейвлетів, виходячи з умов розв'язуваних задач.

Слід зазначити, що майже всі вейвлети не мають аналітичного представлення у вигляді однієї формули, але можуть задаватися ітераційними виразами, які легко обчислюються.

Вейвлети характеризуються своїм часовим і частотним образами. Часовий образ визначається функцією часу $\psi(t)$. А частотний образ визначається її Фур'є-образом $\hat{\psi}(\omega) = F(\omega) \int_{-\infty}^{\infty} \psi(t) e^{-j\omega t} dt$, який задає оригінальну спектра вейвлета.

Розглянемо неперервний сигнал нескінченної тривалості, що описується функцією $s(t)$ - вектором гільбертового простору неперервних функцій однієї змінної: $s(t) \in L^2(\mathbb{R})$, де \mathbb{R} - множина дійсних чисел. Такими функціями можна описати сигнали, які надходять від усіх технічних та біологічних систем.

У просторі існування сигналів можна обрахувати скалярний добуток двох векторів

$$\langle f(t), s(t) \rangle = \int_{-\infty}^{\infty} f(t) s(t) dt.$$

Відомо, що для сигналів гільбертового простору норму можна визначити як

$$\|s(t)\| = \sqrt{\int_{-\infty}^{\infty} |s(t)|^2 dt}.$$

Дійсна функція $\psi(t) \in L^2(\mathbf{R})$, що задовольняє умову допустимості

$$C_{\psi} = \int_0^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty, \quad (2.60)$$

та має одиничну норму

$$\|\Psi(t)\| = 1, \quad (2.61)$$

називається *материнським вейвлетом*. Для заданого масштабу a та зміщення b цей вейвлет породжує набір розтягнутих або стиснутих вейвлет-функцій, кожен з яких можна отримати за правилом

$$\psi(a, b, t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right). \quad (2.62)$$

Якщо було вибрано певний материнський вейвлет і утворено набір функцій (2.62), то функція

$$C(a, b) = \langle s(t), \psi(a, b, t) \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt, \quad a \neq 0 \quad (2.63)$$

є прямим безперервним вейвлет-перетворенням (ПБВП) неперервної функції $s(t)$ відносно материнської функції $\psi(t)$.

Вейвлет-перетворення $C(a, b)$ є неперервною функцією двох дійсних змінних: значень масштабу $a \in \mathbf{R}$, $a \neq 0$, і зміщення $b \in \mathbf{R}$. Параметр b задає положення вейвлетів на осі часу (або по просторовій координаті), а параметр a – його масштаб. Коефіцієнт $\frac{1}{\sqrt{a}}$ забезпечує нормування енергії при зміні масштабу. Параметр масштабу a реально може бути тільки додатним і його не можна брати рівним 0, тобто $a \geq 0$ (умова (8.1) задана саме для такого випадку).

Отже, вейвлет-коефіцієнти визначаються інтегральним значенням скалярного добутку сигналу на вейвлет-функцію $\psi(a, b, t)$ заданого виду, а ПБВП можна розглядати як розкладання сигналу $s(t)$ по всіх можливих зсувах і по розтяганнях (стисканнях) сигналу $s(t)$. При цьому параметри a й b можуть набувати будь-яких значень в межах областей їх визначення. У частотній області малі значення a відповідають високим частотам, а більші – низьким частотам.

При прямому безперервному вейвлет-перетворенні сигнал представляється сукупністю хвильових пакетів – вейвлетів, утворених на основі деякої вихідної (базової) функції $\psi(t)$. Ця сукупність є різною в різних частинах часового інтервалу визначення сигналу і коректується множниками, що мають вид складних часових функцій, та представляє сигнал з тим або іншим ступенем деталізації.

Кількість використуваних при розкладанні сигналу вейвлетів задає *рівень розкладання* або *рівень декомпозиції* сигналу. За нульовий (вищий) рівень вважається сам сигнал, а наступні рівні розкладання утворюють дерево розкладання (декомпозиції) або вейвлет-дерево. Точність представлення сигналу в міру переходу на більш низькі рівні розкладання знижується, але з'являється можливість вейвлет-фільтрації сигналів, усунення шумів із суміші сигнал+шум, стискання сигналів, тобто стає можливою їхня вейвлет-обробка.

Також однією з основних ідей вейвлет-представлення сигналів є розбивка наближення до сигналу на *дві складові* – грубу або *апроксимуючу* й уточнену або, *деталізуючу* – з наступним уточненням цих складових ітераційним методом.

Кожний крок такого уточнення відповідає певному рівню декомпозиції (розкладання) або реставрації (відновлення) сигналу. Ці операції можливі як у часовій, так і в частотній області представлення сигналів вейвлетами. Також в основі ПБВП лежить використання двох безперервних та інтегрованих по всій осі t (або x) функцій:

- вейвлет-функція, яка задає деталізуючі коефіцієнти;
- масштабуюча (апроксимуюча) або скейлінг-функція, яка задає апроксимуючі коефіцієнти.

Слід зазначити кілька важливих обставин. Практично неможливо вести обчислення з безперервними вейвлетами. На практиці доводиться задавати дискретні значення a й b .

Безперервне вейвлет-перетворення має велику надмірність, що веде до невиправдано великих витрат машинного часу, а інколи – до спотворення інтерпретації результатів.

Зворотне безперервне вейвлет-перетворення (ЗБВП) здійснюється по формулі реконструкції в часовій області, яка має цілий ряд форм, що залежать від математичного стилю запису та від визначення областей існування сигналу. Практичний інтерес представляють лише ті формули реконструкції, які реально використовуються в системах комп'ютерної математики при прямому та зворотному безперервних вейвлет-перетвореннях. Наприклад,

$$s(t) = \frac{1}{C_\psi} \int_{R^+ - \{0\}} \int_R \frac{1}{\sqrt{a}} C(a, b) \psi\left(\frac{t-b}{a}\right) \frac{dadb}{a^2}.$$

Тут нижня границя першого інтегралу $R^+ - \{0\}$ означає, що $a \geq 0$.

Вейвлети, що допускають розкладання сигналу і які забезпечують його повну реконструкцію, є ортогональними вейвлетами. Вони є найбільш важливим типом вейвлетів, здатних повноцінно апроксимувати довільні функції та сигнали. При проведенні аналізу сигналів за допомогою неперервного вейвлет-перетворення основним етапом є розрахунок вейвлет-коефіцієнтів. Для розрахунку вейвлет-перетворень сигналів переважно використовується апроксимація неперервного вейвлет-перетворення за допомогою ортогонального діадного вейвлет-перетворення (ОДВП) або застосовується кратномасштабний аналіз (КМА). При цьому необхідно мати формульний вираз для розрахунку значень материнського вейвлету, до того ж, материнська функція має утворювати при масштабуванні ортогональний базис.

Безперервне вейвлет-перетворення вимагає значних обчислювальних витрат. Тому для практичного його застосування необхідна дискретизація значень a і b . Щоб уникнути надмірності вейвлет-перетворення можна задавати дискретні значення a і b на деякій множині $Z = \{\dots -1, 0, 1, 2, \dots\}$, які дорівнюють $a = 2^j$ та $b = k2^j$, де j та k – цілі числа.

Можливі інші способи завдання дискретних значень a і b . Параметр j , іноді позначуваний як m , називається параметром масштабу. Описана дискретизація є найпоширенішою, а сітка дискретизації називається *діадичною*. Відповідно й це перетворення прийнято називати *діадним вейвлет-перетворенням*.

Важливою особливістю діадного вейвлет-перетворення є виключення

перекриття носіїв вейвлетів, тобто усунення надмірності в ході вейвлет-перетворення.

Можливе представлення сигналу на основі підсумовування його грубого представлення з детальними представленнями у його різних місцях. Для цього використовують ортогональні вейвлети. Їх можна створити, ґрунтуючись на представленні простору сигналів V у вигляді вкладених підпросторів V_j , що відрізняються один від одного тільки перемасштабуванням незалежної змінної. Заснований на цьому аналіз називають *кратномасштабним* або *багатомасштабним* аналізом.

Він базується на наступних вихідних передумовах:

- простір сигналів V може бути розбитий на ієрархічно вкладені підпростори V_j , які не перетинаються й об'єднання яких гранично дає норму $L^2(\mathbf{R})$;

- для будь-якої функції $s(t) \in V_j$ її стисла версія належить простору V_{j-1} ;

- існує така функція $\varphi(t) \in V_0$, для якої її зсув $\varphi_{0,j}(t) = \varphi(t - k)$ при $k \in Z$ утворює ортонормований базис простору V_0 .

Оскільки функції $\varphi_{0,k}(t)$ утворюють ортонормований базис простору V_0 , то функції

$$\varphi_{j,k}(t) = 2^{-j/2} \varphi(2^{-j}t - k)$$

утворюють ортонормований базис простору V_j .

Таким чином, функції $\varphi_{j,k}$ називаються масштабуючими саме тому, що вони створюють свої масштабовані версії в просторі сигналу. Їх ще називають материнськими вейвлетами. При цьому сигнал $s(t)$ може бути представлений безліччю послідовних наближень $s_j(t)$ у підпросторах V_j .

У підсумку сума приблизної та детальної компонент дає вихідний сигнал з тим або іншим наближенням. У якості початкового значення часто беруть $j = 0$, але можна починати ітераційний процес наближення й з інших значень j .

Як приклад можна привести передачу зображення, коли спочатку передається грубий образ зображення у вигляді функції $\varphi(t)$, а потім воно уточнюється за допомогою вейвлет-коефіцієнтів і приймає звичний вид.

Причому процес уточнення завжди можна зупинити, якщо, наприклад, досягнута достатня роздільна здатність системи відображення і більше немає необхідності в уточненні зображення або похибка перетворення досяглася рівня похибки машинних розрахунків.

Вибір вейвлетів є досить великим. Широко використовуються близько півтора десятків базових типів вейвлетів і безліч варіантів їх модифікацій. При розв'язанні складних завдань в області обробки сигналів і зображень необхідно застосування хоча б декількох типів вейвлетів з наступним порівнянням результатів і вибором найкращих з них.

До «грубих» вейвлетів або «передвейвлетів» відносяться вейвлети Гаусового типу, вейвлети Морле та вейвлет «мексиканський капелюх». До нескінченних регулярних вейвлетів належать вейвлети Мейєра. До ортогональних вейвлетів з компактним носієм відносяться вейвлети Добеши, симлети та койфлети. До біортогональних парних вейвлетів з компактним носієм відносяться B -сплайнові біортогональні вейвлети.

У системах комп'ютерної математики найчастіше використовуються вейвлети Хаара, Добеши, симлети (симетричні вейвлети Добеши), койфлети (вейвлети Добеши з досить гарною симетрією), Мейєра, дискретна апроксимація вейвлета Мейєра, Гауса, комплексний вейвлет Гауса, Морле, комплексний вейвлет Морле, Шенона, біортогональний, зворотний біортогональний, мексиканський капелюх, частотний B -сплайновий і деякі інші. Варіанти представлення різних вейвлетів за допомогою системи комп'ютерної математики наведені на рис. 2.13 – 2.15.

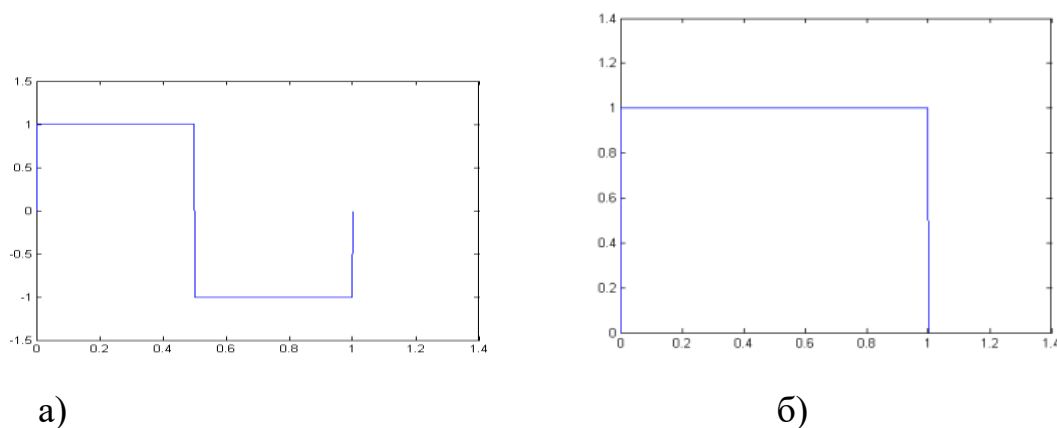
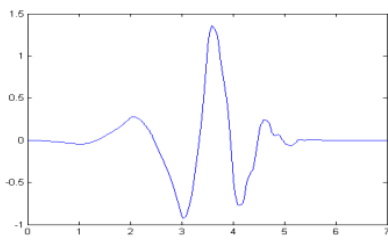
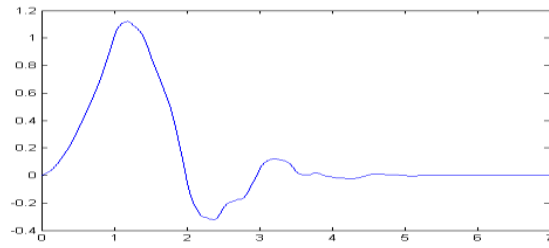


Рис. 2.13. Вейвлет Хаара: а - вейвлет; б - масштабуюча функція

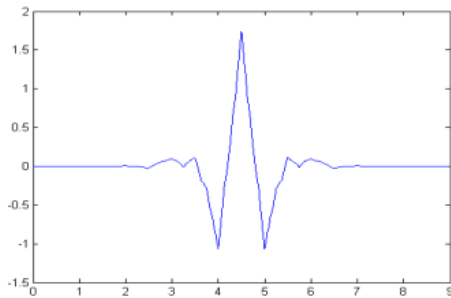


а)

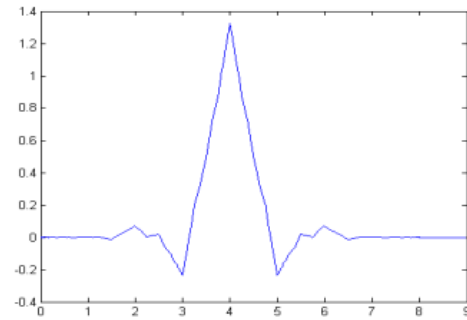


б)

Рис. 2.14. Вейвлет Добеши (db4): а - вейвлет; б - масштабуюча функція



а)



б)

Рис. 2.15. Біортогональні вейвлети (biog4.4): а - вейвлет; б - масштабуюча функція

Питання для самоперевірки

1. Фізичний сенс перетворення Фур'є.
2. В чому відмінність між рядом Фур'є та інтегральним перетворенням Фур'є?
3. Що таке пряме і обернене перетворення Фур'є?
4. Опишіть дискретне перетворення Фур'є.
5. Основні властивості ДПФ.
6. Як залежить протяжність спектру Фур'є від протяжності сигналу?
7. В чому полягає фізичний зміст амплітудного і фазового спектрів?
8. Алгоритм роботи спецпроцесорів ШПФ.
9. Що таке z -перетворення?
10. Які властивості має z -перетворення?
11. Що таке базисна функція вейвлету?
12. В чому полягає ідея вейвлет-перетворення?
13. Опишіть види вагових вікон. Навіщо вони використовуються?

РОЗДІЛ 3. КОРЕЛЯЦІЙНИЙ АНАЛІЗ СИГНАЛІВ ТА ЗОБРАЖЕНЬ

3.1. Основи кореляційного аналізу сигналів та зображень

Кореляційний аналіз разом із спектральним відіграє значну роль в теорії сигналів. Фізичний сенс цього типу аналізу полягає в кількісному вимірюванні міри подібності сигналів. Для цього застосовуються кореляційні функції.

Крім цього, кореляційні властивості є важливою складовою при виборі різновиду сигналу для вирішення тих чи інших завдань.

Вони дають уяву про зміну рівня сигналу на виході пристрою його обробки при відмінності параметрів від очікуваних, що важливо для задач виявлення сигналів та вимірювання їх параметрів на тлі завад.

Кореляційні функції можна розраховувати як для двох різних сигналів, так і для одного того ж сигналу, при різних його положеннях на осі часу. В другому випадку функція називається автокореляційною.

3.2. Автокореляційна функція

Завдання порівняння сигналу з його зсунутою в часі копією, є типовим для локаційних методів вимірювання дальності до об'єкта (радіолокація, гідролокація). В цьому випадку затриманий сигнал, що відбився від опромінюваного об'єкта, має практично таку ж форму, що й зондуєчий сигнал. В пристрої вимірювання часу затримки відбитого сигналу виконується порівняння цього сигналу з копіями зразкового (зондуєчого) сигналу зсунутими на фіксовані відрізки часу. В результаті порівняння в кожному з каналів двох сигналів обирається канал, на виході якого схожість сигналів є найкращою.

За номером цього каналу отримується значення часу затримки відбитого сигналу і, відповідно, дальності до об'єкта.

Якість роботи такого пристрою буде залежати від того, якою мірою будуть відрізнятись між собою сигнали при порівняно невеликих відносних зсувах. Для кількісної оцінки такої відмінності застосовується автокореляційна функція сигналу (АКФ, correlation function).

Автокореляційна функція детермінованого сигналу $s(t)$ з кінцевою енергією є функцією часу τ відносного зсуву двох копій сигналу і є скалярним добутком

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t)s(t - \tau)dt. \quad (3.1)$$

Автокореляційна функція є мірою схожості між сигналом і його зсунутою копією – чим більше значення АКФ, тим ця схожість є більшою.

АКФ має наступні властивості:

1. Значення АКФ при $\tau = 0$ дорівнює енергії сигналу, тобто інтегралу від його квадрата

$$B_s(0) = \int_{-\infty}^{\infty} s(t)s(t)dt = \int_{-\infty}^{\infty} s^2(t)dt = E.$$

2. АКФ є парною функцією свого аргументу τ

$$B_s(\tau) = B_s(-\tau).$$

3. Значення АКФ при $\tau = 0$ є максимально можливим значенням: $B_s(\tau) \leq B_s(0)$.

4. Із зростанням абсолютного значення τ АКФ сигналу з кінцевою енергією затухає $\lim_{|\tau| \rightarrow \infty} B_s(\tau) = 0$.

5. Якщо сигнал $s(t)$ не містить δ -функцій, то його АКФ не може мати розривів (тобто обов'язково буде неперервною функцією).

Розглянемо обчислення АКФ на прикладі прямокутного імпульсу тривалістю T , який розпочинається в нульовий момент часу (рис. 3.1):

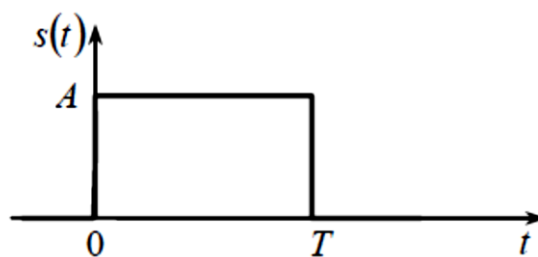


Рис. 3.1. Прямокутний імпульс

$$\text{при } -T \leq \tau \leq 0: B_s(\tau) = \int_0^{T+\tau} A^2 dt = A^2(T + \tau);$$

$$\text{при } 0 \leq \tau \leq T: B_s(\tau) = \int_{\tau}^T A^2 dt = A^2(T - \tau);$$

$$\text{при } |\tau| > T(0): B_s(0) = 0.$$

Остаточно

$$B_s(\tau) = \begin{cases} A^2(T - |\tau|), & \text{якщо } |\tau| \leq T; \\ 0, & \text{якщо } |\tau| > T. \end{cases}$$

Графік АКФ прямокутного імпульсу показано на рис. 3.2.

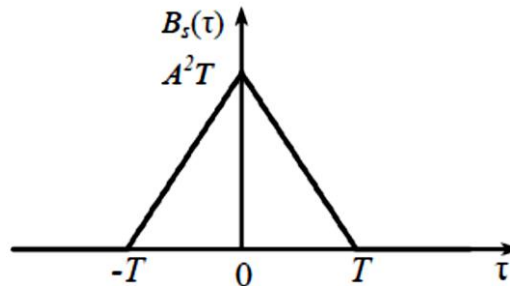


Рис. 3.2. Кореляційна функція прямокутного імпульсу

У разі періодичного сигналу (і взагалі будь-якого сигналу з нескінченною енергією) скористатися наведеним вище визначенням АКФ не вдасться. Тому АКФ періодичного сигналу з періодом T обчислюють, усереднюючи добуток зсунутих копій в межах одного періоду:

$$B_s(\tau) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t - \tau)dt. \quad (3.2)$$

Набір властивостей *періодичного сигналу* АКФ (3.2) дещо змінюється порівняно з (3.1):

1. При $\tau = 0$ її значення дорівнює не енергії, а середній потужності аналізованого сигналу

$$B_s(0) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s^2(t)dt.$$

2. Властивість парності зберігається $B_s(\tau) = B_s(-\tau)$.

3. Значення АКФ при $\tau=0$, як і раніше, залишається максимально можливим $B_s(\tau) \leq B_s(0)$.

4. АКФ періодичного сигналу є періодичною функцією з тим же періодом, що й сам сигнал: $B_s(\tau + T) = B_s(\tau)$.

5. Якщо сигнал не містить дельта-функцій, його АКФ буде неперервною функцією.

Як приклад, обчислимо АКФ гармонічного сигналу

$$s(t) = A \cos(\omega_0 t + \varphi_0).$$

Період такого сигналу дорівнює $2\pi/\omega_0$. Тому кореляційний інтеграл

$$B_s(\tau) = \frac{\omega_0}{2\pi} \int_{-\frac{\pi}{\omega_0}}^{\frac{\pi}{\omega_0}} A \cos(\omega_0 t + \varphi_0) A \cos(\omega_0(t - \tau) + \varphi_0) dt = \frac{A^2}{2} \cos(\omega_0 \tau).$$

Як видно, АКФ гармонічного сигналу також є гармонічною функцією. Важливо також, що отриманий результат не залежить від початкової фази (параметр φ_0 до отриманого виразу не увійшов). Це прояв загальної властивості всіх кореляційних функцій.

3.3. Функція взаємної кореляції

Якщо АКФ показує міру схожості між зсунутими копіями одного й того ж сигналу, то взаємна кореляційна функція (ВКФ, cross-correlation function) дозволяє обчислювати аналогічну величину для зсунутих екземплярів двох різних сигналів.

Загальний вигляд формули кореляційної функції зберігається, але під інтегралом стоїть добуток двох різних сигналів, один з яких є затриманим на певний час τ

$$B_{12}(\tau) = \int_{-\infty}^{\infty} s_1(t) s_2(t - \tau) dt.$$

Очевидно, що АКФ є окремим випадком ВКФ, коли обидва сигнали є однаковими: $s_1(t) = s_2(t) = s(t)$.

Як приклад, обчислимо ВКФ прямокутного і трикутного імпульсів (рис. 3.1, рис. 3.3):

$$s_1(t) = \begin{cases} A, & 0 \leq t \leq T, \\ 0, & t < 0, t > T; \end{cases} \quad s_2(t) = \begin{cases} A \frac{t}{T}, & 0 \leq t \leq T, \\ 0, & t < 0, t > T. \end{cases}$$

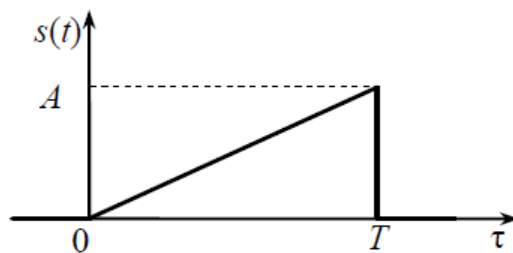


Рис. 3.3. Трикутний імпульс

Маємо три випадки

- при $0 \leq \tau \leq T$

$$B_{12}(\tau) = \int_{\tau}^T A^2 \frac{t - \tau}{T} dt = \frac{A^2}{2T} (T - \tau)^2;$$

- при $-T \leq \tau \leq 0$

$$B_{12}(\tau) = \int_{\tau}^T A^2 \frac{t - \tau}{T} dt = \frac{A^2}{2T} (T^2 - \tau^2);$$

- при $|\tau| > T$

$$B_{12}(\tau) = 0.$$

Остаточно можна записати

$$B_{12}(\tau) = \begin{cases} \frac{A^2}{2T} (T - \tau)^2, & 0 \leq \tau \leq T, \\ \frac{A^2}{2T} (T^2 - \tau^2), & -T \leq \tau \leq 0, \\ 0, & |\tau| > T. \end{cases}$$

Графік отриманої ВКФ показано на рис. 3.4.

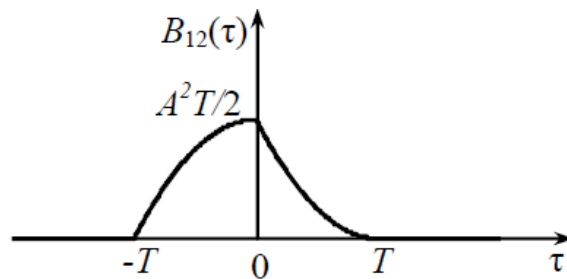


Рис. 3.4. ВКФ прямокутного і трикутного імпульсів

Властивості ВКФ дещо відрізняються від властивостей АКФ:

1. $|B_{12}(\tau)| \leq \sqrt{E_1 E_2}$, де E_1, E_2 – енергії сигналів.
2. $B_{12}(-\tau) = B_{21}(\tau)$, тобто зміна знаку τ рівнозначна взаємній перестановці сигналів.
3. Значення ВКФ при $\tau = 0$ не є особливим. Її максимум може бути в будь-якій точці на осі τ .
4. Із зростанням абсолютного значення τ ВКФ сигналів з кінцевою

енергією затухає, тобто $\lim_{|\tau| \rightarrow \infty} B_{12}(\tau) = 0$.

5. Якщо сигнали $s_1(t)$ і $s_2(t)$ не містять особливостей у вигляді дельта-функцій, їх ВКФ не матиме розривів (тобто обов'язково буде неперервною функцією).

Для періодичних сигналів поняття ВКФ зазвичай не застосовується, хоча воно може бути введено у разі, якщо сигнали $s_1(t)$ і $s_2(t)$ мають однаковий період.

3.4. Зв'язок між кореляційними функціями і спектрами сигналів

Оскільки кореляційні функції, як і спектри є інтегральними перетвореннями аналізованих сигналів, логічно припустити, що ці характеристики є пов'язаними між собою. Для виявлення цього зв'язку піддамо взаємну кореляційну функцію перетворенню Фур'є, вважаючи, що сигнали $s_1(t)$ і $s_2(t)$ мають спектральні функції $\dot{S}_1(\omega)$ і $\dot{S}_2(\omega)$

$$\begin{aligned} \int_{-\infty}^{\infty} B_{12}(\tau) \exp(-j\omega\tau) d\tau &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_1(t)s_2(t-\tau) \exp(-j\omega\tau) dt d\tau = \\ &= \int_{-\infty}^{\infty} s_1(t) \exp(-j\omega t) \int_{-\infty}^{\infty} s_2[-(\tau-t)] \exp(-j\omega(\tau-t)) d(\tau-t) dt = \dot{S}_1(\omega)\dot{S}_2^*(\omega). \end{aligned} \quad (3.3)$$

При обчисленні другого інтегралу враховано властивість 3 перетворення Фур'є (зміна масштабу осі часу) у випадку $a = -1$ (див. підрозділ 2.2).

Таким чином, ВКФ пов'язана перетворенням Фур'є з так званим взаємним спектром сигналів, який для сигналів $s_1(t)$ і $s_2(t)$ є добутком їх спектральних функцій, одна з яких піддана комплексному спряженню

$$\dot{S}_{12}(\omega) = \dot{S}_1(\omega)\dot{S}_2^*(\omega). \quad (3.4)$$

Звідси можна зробити дуже важливий висновок: якщо спектри сигналів не перекриваються, то їх взаємний спектр дорівнює нулю на усіх частотах, а це означає, що дорівнює нулю і їх ВКФ при будь-яких часових зсувах τ .

Таким чином, сигнали із спектрами, що не перекриваються, є некорельованими. Приймавши $s_1(t) = s_2(t) = s(t)$, отримуємо аналогічний результат для АКФ

$$\int_{-\infty}^{\infty} B_s(\tau) e^{-j\omega\tau} d\tau = \dot{S}(\omega)\dot{S}^*(\omega) = |\dot{S}(\omega)|^2. \quad (3.5)$$

Отже, АКФ сигналу пов'язана перетворенням Фур'є з квадратом модуля спектральної щільності (з енергетичним спектром сигналу).

Таким чином, обернене перетворення Фур'є енергетичного спектру сигналу дозволяє обчислити його АКФ

$$B_s(\tau) = \frac{1}{2\pi j} \int_{-\infty}^{\infty} |\dot{S}(\omega)|^2 e^{j\omega\tau} d\omega. \quad (3.6)$$

Розглянемо приклад обчислення АКФ сигналу, який має рівномірний і обмежений за частотою енергетичний спектр

$$W_u(\omega) = \begin{cases} W_0, & -\omega_B \leq \omega \leq \omega_B, \\ 0, & |\omega| > \omega_B, \end{cases}$$

де ω_B – верхня гранична частота спектру.

За формулою (3.6) маємо АКФ

$$B_u(\tau) = \frac{W_0}{2\pi} \int_{-\omega_B}^{\omega_B} e^{j\omega\tau} d\omega = \frac{W_0}{\pi} \int_0^{\omega_B} \cos \omega\tau d\omega = \frac{W_0 \omega_B}{\pi} \frac{\sin \omega_B \tau}{\omega_B \tau}.$$

Отримана АКФ сигналу з прямокутним спектром подібно до спектру прямокутного імпульсу має пелюстковий характер. Ширину головної пелюстки часто приймають за інтервал кореляції, який отримуємо з співвідношення $\omega_B \tau_k = \pi$, тобто $\tau_k = \pi / \omega_B = 1 / (2f_B)$.

Кореляційні властивості сигналу, таким чином, можуть бути оцінені за його енергетичним спектром. Чим ширше спектр сигналу, тим вужче головна пелюстка його АКФ і тим він краще з точки зору вимірювання, наприклад, моменту появи відбитого сигналу в процесі виявлення цілі при застосуванні локаційних методів. Окрім цієї властивості, важливою кореляційною властивістю сигналу є й рівень бічних пелюсток сигналу. Для виявлення слабких сигналів поруч з більш потужними потрібно обирати такі сигнали, які забезпечують порівняно низький рівень бічних пелюсток відносно головної пелюстки. З (3.5) також видно, що АКФ сигналу не залежить від його фазового спектру.

Отже, сигнали, амплітудні спектри яких однакові, а фазові різняться, матимуть однакову АКФ. Тобто маємо ще один наслідок – за АКФ неможливо відновити вихідний сигнал (із-за втрати інформації про фазу).

Взаємний спектр сигналів $s_1(t)$ і $s_2(t)$ є прямим перетворенням Фур'є їх функції взаємної кореляції (3.3). Відповідно, якщо відомий взаємний

спектр сигналів, їх ВКФ може бути обчислена за формулою оберненого перетворення Фур'є

$$B_{12}(\tau) = \frac{1}{2\pi} \int_{-\omega_B}^{\omega_B} \dot{S}_{12}(\omega) e^{j\omega\tau} d\omega. \quad (3.7)$$

Підстановка значення $\tau=0$ в (3.4) з розкриттям виразів для ВКФ і взаємного спектру дає рівність

$$\int_{-\infty}^{\infty} S_1(t) S_2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{S}_1(\omega) \dot{S}_2^*(\omega) d\omega. \quad (3.8)$$

Вираз (3.8) може бути поданий у більш загальному вигляді

$$(s_1, s_2) = \frac{1}{2\pi} (\dot{S}_1, \dot{S}_2). \quad (3.9)$$

Рівність (3.9) називається *узагальненою формулою Релея*. Вона є справедливою також у випадку, коли сигнал є комплексним.

Узагальнена формула Релея може трактуватись наступним чином: *скалярний добуток двох сигналів з точністю до коефіцієнта дорівнює скалярному добутку їх спектральних щільностей*.

Якщо сигнали однакові $s_1(t) = s_2(t) = s(t)$ з (3.9) отримаємо співвідношення для обчислення енергії сигналу як в часовій, так і в частотній області, яке є різновидом *рівності Парсеваля* для фінітних сигналів

$$E = \int_{-\infty}^{\infty} s^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\dot{S}(\omega)|^2 d\omega. \quad (3.10)$$

Питання для самоперевірки

1. В чому полягає фізичний сенс автокореляційної функції?
2. В чому полягає фізичний сенс функції взаємної кореляції?
3. Зв'язок між кореляційними функціями і спектрами.
4. В чому сенс формули Релея?
5. Як можна оцінити кореляційні властивості сигналу за його енергетичним спектром?

РОЗДІЛ 4. АНАЛІЗ ВИПАДКОВИХ СИГНАЛІВ

4.1. Характеристики випадкових сигналів

Випадковою називають таку величину, яка залежно від випадку набуває того чи іншого чисельного значення. Оскільки закономірності у появі цих значень немає, аналіз таких величин може виконуватись лише методами теорії ймовірності.

Найчастіше зустрічаються випадкові величини двох типів: дискретні та безперервні. Випадкова величина дискретного типу характерна тим, що може приймати лише кінцеві, ізольовані значення і не може набувати проміжних значень. Прикладом величин цього типу може бути число непридатних деталей партії. Характерною особливістю випадкової величини безперервного типу і те, що ця величина може набувати будь-які числові значення в межах певного інтервалу, тобто може мати нескінченне число можливих значень в межах інтервалу. Прикладом випадкових величин цього типу можуть бути, наприклад, дійсні розміри деталей у партії.

На відміну від детермінованих сигналів, форма яких відома, миттєві значення випадкових сигналів заздалегідь не відомі і можуть бути передбачені лише з деякою ймовірністю, яка менша за одиницю. Характеристики таких сигналів є ймовірнісними.

У технічній літературі терміни «випадковий сигнал» і «випадковий процес» зазвичай використовуються як синоніми. Застосовується також термін «стохастичний процес».

Випадковий процес (сигнал) $X(t)$ – це випадкова функція, миттєве значення якої в будь-який момент часу $t = t_0$ є випадковою величиною $X(t_0)$. Або інакше – це сукупність (*ансамбль*) $\{x_k(t)\}$, $k=1, 2, 3, \dots$ всіх можливих вибіркової функцій (*реалізацій*), які породжуються певним випадковим явищем.

Одна з цих реалізацій, наприклад $x_k(t)$, стає повністю відомою після реєстрації (прийому) сигналу. Вона із випадкової перетворюється на детерміновану функцію часу. Випадкові сигнали часто можуть бути доволі простими сигналами, в яких випадковим є один або кілька параметрів цього

сигналу. Такі випадкові сигнали називають квазідетермінованими.

Для аналізу властивостей і характеристик випадкового сигналу, його перетворень в системах обробки сигналів потрібно задати його математичну модель. Така модель може описувати можливі реалізації випадкового сигналу разом із вказівкою відносної частоти їх появи.

Випадковий процес може залежати не лише від одного аргументу – часу, але й від кількох, наприклад часу і просторових координат (приклад – швидкість вітру). Такі випадкові процеси називають *випадковими полями*. Випадкові поля можуть бути скалярними і векторними. Тому додатково до введеної на початку курсу класифікації сигналів випадкові процеси можуть поділятися на:

- *скалярний* $X(t)$ – випадковий процес, область значень якого є множиною в просторі дійсних чисел;
- *векторний* $X(t)$ – випадковий процес, область значень якого є множиною у відповідному координатному просторі;
- *скалярне випадкове поле* $X(r, t)$ – випадковий процес, область значень якого є множиною з дійсних чисел у відповідному координатному просторі;
- *векторне випадкове поле* $X(r, t)$ – випадковий процес, що представляється у вигляді компонент, які є скалярними полями;

Розглянемо як приклад модель гармонічного сигналу з випадковою початковою фазою. Така модель скалярного випадкового сигналу використовується у багатьох практичних задачах. Реалізації сигналу є гармонічними коливаннями з відомою амплітудою і частотою, але з випадковою початковою фазою. Кожна з них може бути записана як

$$s(t) = A \cos(\omega_0 t + \varphi),$$

де A – амплітуда (детермінована), ω_0 – частота (детермінована) і φ – випадкова початкова фаза, яка в більшості практично цікавих випадків може вважатися рівномірно розподіленою на інтервалі $(0 - 2\pi)$.

Графіки декількох реалізацій цього випадкового сигналу косинусоїди, що зміщені одна відносно одної за часом показано на рис. 4.1. Конкретний вигляд реалізації на рис. 4.1 визначається значенням лише однієї випадкової величини – початкової фази. Тобто маємо приклад квазідетермінованого випадкового сигналу.

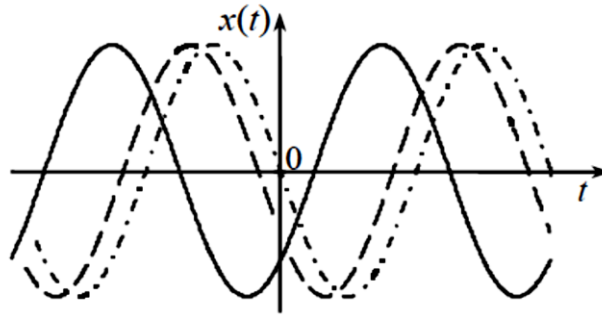


Рис. 4.1. Кілька реалізацій гармонічного сигналу з випадковою початковою фазою

Для того щоб знати випадкову величину, насамперед необхідно знати ті значення, які вона може набувати, і з якою ймовірністю вона набуває цих значень. Ймовірнісною характеристикою випадкової величини є функція розподілу випадкових величин, що показує, з якою ймовірністю випадкова величина набуває тих чи інших числових значень. Таку функцію для величин дискретного типу називають функцією розподілу ймовірностей.

Функція розподілу $P(x_i)$ показує ймовірність появи кожного з можливих значень x_i випадкової величини x .

На відміну від дискретних величин ймовірність того, що випадкова величина безперервного типу прийме яесь певне значення X , дорівнює нулю, так як кількість можливих значень є нескінченною. Тому для безперервних випадкових величин обчислюють ймовірність того, що можливе значення величини x знаходиться в заданому інтервалі, а не ймовірність того, що величина x прийме певне значення X .

Як імовірнісну характеристику випадкової величини в цьому випадку використовують поняття *щільності ймовірності*. Щільністю ймовірності безперервної випадкової величини називають границю відношення ймовірності того, що можливі значення величини x знаходяться в інтервалі від X до $X+\Delta x$ до довжини інтервалу Δx , коли Δx наближається до нуля

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(X \leq x < X + \Delta x)}{\Delta x} \quad (4.1)$$

Функцію $f(x)$ називають також диференціальним законом розподілу чи диференціальною функцією розподілу.

Щільність розподілу є розмірною функцією:

$$\dim f(x) = \dim \frac{1}{x}.$$

З точністю до нескінченно малих вищого порядку щодо dx величина $f(x)dx$ виражає ймовірність того, що випадкова величина x прийме якийсь із можливих значень інтервалу $[X, X+dx]$, тобто.

$$P(X < x < X + dx) = f(x)dx, \quad (4.2)$$

де dx – диференціал x .

Повна площа під кривою $f(x)$ дорівнює ймовірності появи будь-якого з можливих значень x тобто дорівнює 1.

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Істотним недоліком щільності ймовірності є неможливість її застосування до випадкових величин дискретного типу.

Загальним поняттям, яке можна застосовувати до випадкових величин як дискретного, так і безперервного типу, є інтегральна функція розподілу $F(x)$, яка визначає ймовірність того, що випадкова величина x буде меншою, ніж деяке кінцеве число X

$$F(x) = P(x < X). \quad (4.3)$$

Функція розподілу є неспадною функцією, визначеною так, що $F(-\infty)=0$, а $F(+\infty)=1$ (рис. 4.2).

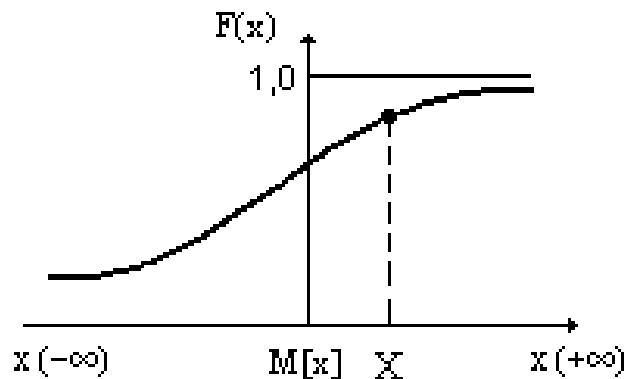


Рис. 4.2.

Інтегральна

функція розподілу $F(x) = P(x < X)$

Для дискретних величин

$$F(x) = P(x < X) = \sum_{x_j} P(x_j).$$

Для безперервних величин

$$F(x) = P(x < X) = \int_{-\infty}^x f(x) dx .$$

Реальні процеси у різних системах моделюються декількома типами законів розподілу випадкових величин: рівної ймовірності, закону Сімпсона, Релея, Райса, хі-квадрат розподіл, нормальний розподіл тощо.

У практиці точних вимірювань найчастіше мають справу з нормальним розподілом результатів. Такий закон діє, коли на процес впливає одночасно велика кількість незалежних та однорідних за ступенем свого впливу випадкових факторів. При цьому вплив кожного окремого фактору є набагато меншим, ніж вплив всієї сукупності.

На практиці замість повного статистичного опису властивостей сукупності випадкових величин x часто обмежуються лише зазначенням деяких часткових характеристик цієї сукупності - моментів розподілу - початкових та центральних.

Початковим моментом k -го порядку випадкової величини x називають математичне очікування k -того ступеня

$$M[x]^k = \int_{-\infty}^{+\infty} x^k f(x) dx. \quad (4.4)$$

Серед початкових моментів найважливішим є перший, який зазвичай називають просто математичним очікуванням

$$M[x] = m_x = \int_{-\infty}^{+\infty} x f(x) dx. \quad (4.5)$$

Перший початковий момент характеризує положення центру розподілу – точки, до якої тяжіє сукупність значень випадкової величини (тобто значення координати центру тяжіння фігури, утвореної віссю абсцис та кривої розподілу цієї випадкової величини).

Центральним моментом k -го порядку випадкової величини x називають математичне очікування k -ого ступеня її відхилення від середнього значення

$$\eta_k(x) = M[x - M[x]]^k = \int_{-\infty}^{+\infty} (x - M[x])^k f(x) dx . \quad (4.6)$$

Перший центральний момент завжди дорівнює 0:

$$\eta_1(x) = M[x - M[x]] = M[x] - M[x] = 0.$$

Другий центральний момент характеризує розсіювання випадкової величини x , тобто розкид її значень щодо центру групування та називається *дисперсією* $D[x]$

$$\eta_2(x) = M[x - M[x]]^2 = D[x] = \int_{-\infty}^{+\infty} (x - M[x])^2 f(x) dx. \quad (4.7)$$

Розмірність дисперсії відрізняється від розмірності випадкової величини x . Тому замість дисперсії часто застосовують додатній корінь із неї, який називають *середнім квадратичним відхиленням* (СКВ)

$$\text{СКВ} = \sigma[x] = \sigma_x = +\sqrt{D[x]}. \quad (4.8)$$

Вказані засоби статистичного опису властивостей випадкових величин відносяться до їхньої нескінченної сукупності.

Оскільки на практиці число спостережень n значень величини x є обмеженим, то за даними такої випадкової вибірки x_1, x_2, \dots, x_n визначити істинні значення невідомих параметрів розподілу m_x і σ_x неможливо. Замість них визначається лише їх статистичні оцінки, які, будучи функціями членів вибірки, відхиляються від істинних значень відповідних параметрів.

Оцінкою істинного значення математичного очікування випадкової величини x є середнє арифметичне вибірки

$$M[x] \approx \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (4.9)$$

При великому числі n значень випадкової величини x у вибірці, оцінку її СКВ можна обчислити за формулою

$$\sigma_x \approx \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (4.10)$$

Для малої кількості n застосовують формулу Бесселя, у яку вводиться поправний множник зменшення зміщення

$$\sigma_x \approx \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}. \quad (4.11)$$

4.2. Опис випадкових функцій

Конкретний вид, який приймає випадкова функція в результаті одного

дослід/вимірювання, називається її реалізацією. В межах отриманої реалізації функція є відомою, тому її можна обробляти як звичайну відому функцію.

Першою характеристикою випадкової функції аналогічно розглянутому раніше випадку випадкової величини є її середнє значення (математичне очікування). Середні значення позначають рискою над символом величини, наприклад $\overline{U(t)}$, а математичне очікування позначають як $M[U(t)]$.

Визначення середнього значення випадкової функції може бути зроблено двома способами: або шляхом усереднення в часі, або усереднення за сукупністю реалізацій.

Першим способом випадкову величину усереднюють у часі в межах однієї реалізації (рис. 4.3, а).

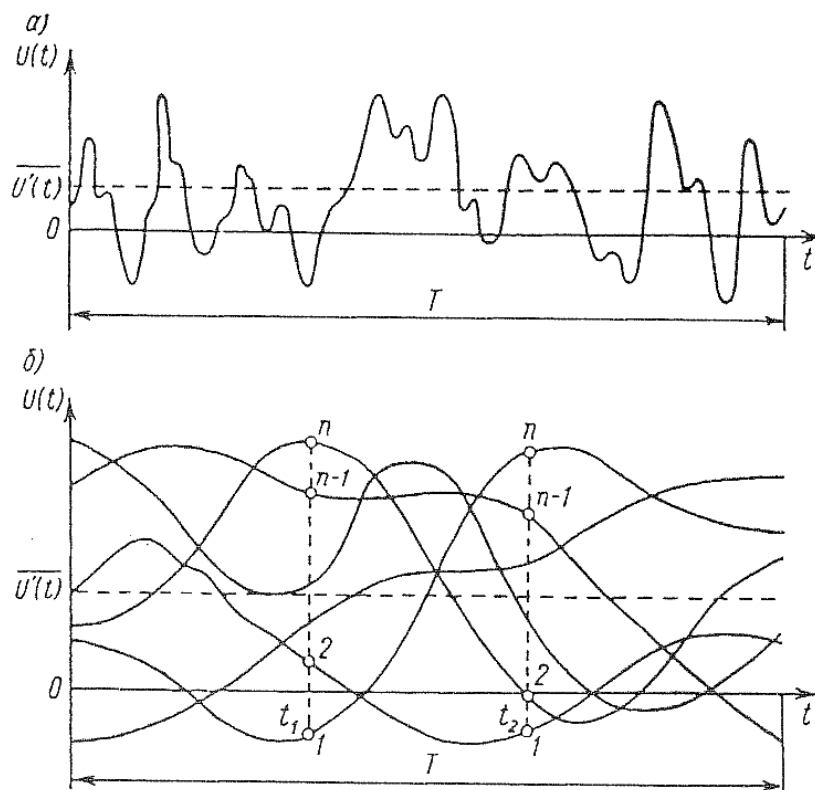


Рис. 4.3. Реалізація випадкової функції:

а – одиночна реалізація; б - сукупність реалізацій

Щоб зробити таке усереднення, розіб'ємо час реалізації T на велике число n рівних інтервалів Δt , як показано на рис. 4.4.

Вважаючи значення $U(t_i)$ постійним у межах інтервалу Δt , здійснимо

відліки і знайдемо суму цих значень. Потім суму розділимо на число $n = \frac{T}{\Delta t}$.
Отримаємо середнє у часі значення однієї реалізації (рис. 4.3, а)

$$\overline{U'(t)} = \frac{1}{n} \sum_{i=1}^n U(t_i) = \frac{1}{T} \sum_{i=1}^n U(t_i) \Delta t . \quad (4.12)$$

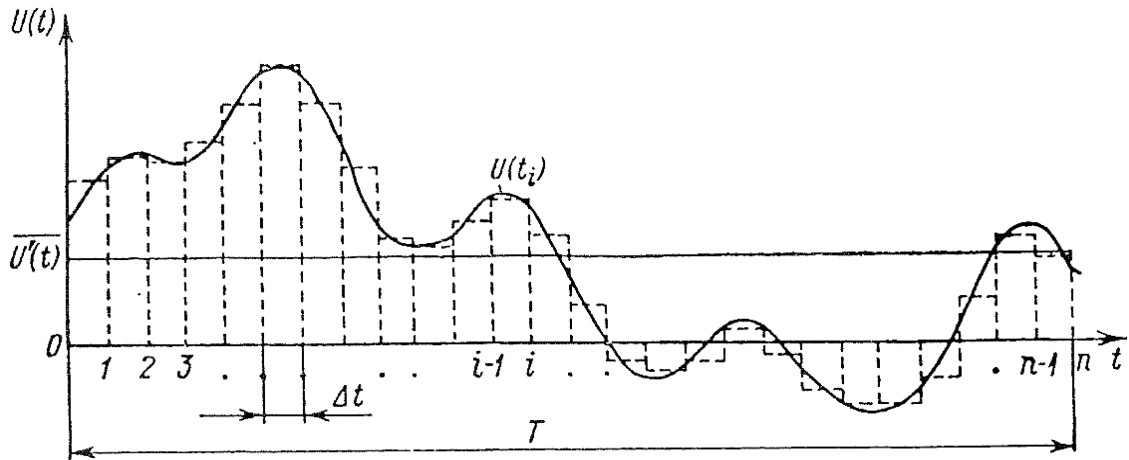


Рис. 4.4. Розбиття часу реалізації при визначенні середнього значення випадкової функції

При збільшенні числа n та відповідному зменшенні інтервалів Δt у граничному наближенні отримаємо

$$\overline{U'(t)} = \frac{1}{T} \int_0^T U(t) dt.$$

Штрих означає, що величина $\overline{U'(t)}$ є наближеним середнім значенням. Точне середнє у часі значення випадкової функції дорівнює

$$\overline{U(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T U(t) dt. \quad (4.13)$$

За другим способом для усереднення використовується велика кількість реалізацій n при одному фіксованому значенні аргументу t (на рис. 4.3, б - t_1). Визначимо n значень випадкової функції, які відповідають моменту часу t_1 , і позначимо їх як $U_1(t_1), U_2(t_1), \dots, U_n(t_1)$. Потім знайдемо середнє значення цих величин, склавши їх і розділивши на n . Отримаємо

$$\overline{U'(t_1)} = \frac{1}{n} \sum_{i=1}^n U_i(t_1). \quad (4.14)$$

Якщо кількість реалізацій n є настільки великою, що її збільшення не змінює помітно величини $\overline{U'(t_1)}$, можна вважати цю величину близькою до середнього значення, яке визначається виразом

$$\overline{U(t_1)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n U_i(t_1). \quad (4.15)$$

Визначення середнього значення можна повторити для іншого фіксованого часу t_2 і отримати величину $\overline{U'(t_2)}$, яка є близькою до точного середнього значення

$$\overline{U(t_2)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n U_i(t_2).$$

Значення $\overline{U'(t_1)}$ і $\overline{U'(t_2)}$ в загальному випадку можуть відрізнятися одне від одного. Однак у найцікавіших для практики випадках ці величини виявляються практично рівними між собою і їх середнє значення є однаковим для всіх моментів часу. Випадкові функції, що мають цю властивість, називаються стаціонарними. Їх середнє значення є постійним числом: $\overline{U'(t_1)} = \overline{U'(t_1 + t)} = \overline{U(t)}$, де t – довільний час.

Отже *стаціонарні випадкові процеси* – це випадкові процеси, статистичні характеристики яких однакові в усіх часових перерізах.

Випадковий процес є *стро́го стаціонарним* (стаціонарним у вузькому сенсі), якщо його багатовимірна щільність ймовірності $p(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n)$ довільної розмірності n не змінюється при одночасному зсувові усіх часових перерізів t_1, t_2, \dots, t_n уздовж осі часу на будь-яку однакову величину τ :

$$p(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = p(x_1, x_2, \dots, x_n, t_1 + \tau, t_2 + \tau, \dots, t_n + \tau).$$

Якщо ж обмежитись вимогами, щоб від часового зсуву не залежали лише одновимірна і двовимірна щільності ймовірності, то такий випадковий процес буде *стаціонарний в широкому сенсі*.

Подальше спрощення аналізу випадкових процесів досягається при виконанні умови ергодичності процесу.

Стаціонарний випадковий процес називається *ергодичним*, якщо при визначенні будь-яких його моментних функцій статистичне усереднення по ансамблю реалізацій еквівалентне усереднюванню за часом однієї, теоретично нескінченно довгої, реалізації $x(t)$.

Позначивши усереднення кутовими дужками, запишемо вирази для обчислення важливих статистичних характеристик ергодичного процесу за його єдиною реалізацією $x(t)$ (ергодичний процес обов'язково є стаціонарним, але не навпаки)

$$m_x(x) = \langle x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt.$$

$$D_x = \langle [x(t) - m_x]^2 \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) dt - m_x^2.$$

$$R_x(\tau) = \langle [x(t) - m_x][x(t-\tau) - m_x] \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau) dt - m_x^2.$$

Математичне очікування ергодичного випадкового процесу дорівнює постійній складовій будь-якої його реалізації.

Дисперсією випадкової функції називається середнє значення квадрата відхилення випадкової функції від її середнього значення

$$D[U(t)] = \overline{\Delta U^2(t)} = \overline{[U(t) - \overline{U(t)}]^2}. \quad (4.16)$$

Тут $\Delta U(t) = U(t) - \overline{U(t)}$ – відхилення від середнього значення.

На рис. 4.5, а) і б) зображені дві випадкові функції з однаковим середнім значенням, але які помітно відрізняються відхиленнями. Дисперсія характеризує величину цих відхилень.

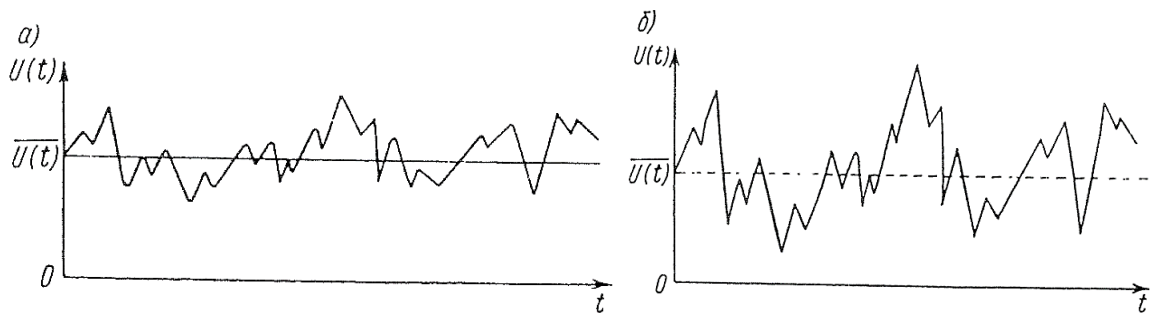


Рис. 4.5. Різні випадки реалізації випадкової функції: а, б - сигнали з різною дисперсією

Необхідність квадратичної залежності у визначенні дисперсії викликана тим, що середнє значення відхилень виявиться близьким або рівним нулю і не дасть уявлення про величину відхилень. Додатні та від'ємні відхилення є однаково ймовірними і в сумі дали б нуль. Квадрат відхилення завжди є додатнім. Тому з допомогою суми можна врахувати всі відхилення.

Дисперсію можна визначити за великою кількістю реалізацій n аналогічно тому, як це робилося щодо середнього значення. З огляду на стаціонарності аналізованих процесів це можна зробити для будь-якого

моменту часу. Отже маємо

$$D_n[U(t)] = \overline{\Delta U^2(t)} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [U_i(t) - \overline{U(t)}]^2. \quad (4.17)$$

Для стаціонарних випадкових процесів дисперсія є постійним числом. Дисперсію можна також визначити як значення у часі для однієї реалізації

$$D_t[U(t)] = \overline{\Delta U^2(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T [U(t) - \overline{U(t)}]^2 dt. \quad (4.18)$$

Слід враховувати, що реалізація $U(t)$ може мати різний сенс. Для процесів в електричних ланцюгах під випадковою функцією $U(t)$ зазвичай розуміють напругу шумів.

В електричних ланцюгах легко відокремити середнє значення (постійну складову) за допомогою роздільного RC-фільтра та отримати центровані випадкові функції струму та напруги. Тому надалі часто приймається $\overline{U(t)} = 0$ і розглядається сигнал без постійної складової (рис. 4.6).

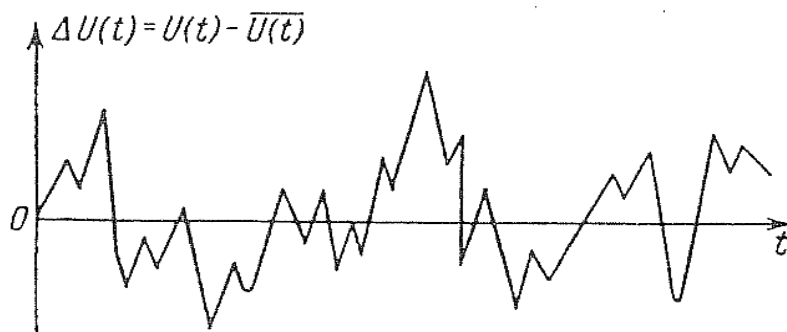


Рис. 4.6. Центрована випадкова функція

Середнє значення потужності, що виділяється в опорі R шумовим струмом або напругою при $\overline{U(t)} = 0$, складає

$$P = \frac{\overline{\Delta U^2(t)}}{R} = \overline{\Delta I^2(t)} R. \quad (4.19)$$

Оскільки для змінного струму потужність дорівнює $\frac{U_D^2}{R}$ та $I_D^2 R$, де U_D та I_D - діючі значення, можна вважати, що дисперсії $\overline{\Delta U^2(t)}$ та $\overline{\Delta I^2(t)}$ дорівнюють квадратам діючих значень $\frac{U_D^2}{R}$ і $I_D^2 R$. Тоді діючі значення шумового струму та напруги рівні їх середньоквадратичним відхиленням:

$$U_D = \sqrt{\overline{\Delta U^2(t)}}; \quad I_D = \sqrt{\overline{\Delta I^2(t)}}. \quad (4.20)$$

Дисперсія ергодичного випадкового процесу має наочний фізичний зміст потужності флуктуаційної складової.

Третьою усередненою величиною, що характеризує випадкові функції, є *функція кореляції*. Ще одна характеристика випадкового процесу необхідна тому, що середнє значення та дисперсія не описують його досить повно.

На рис. 4.7 наведено сукупності реалізацій двох випадкових величин. Дисперсія та математичне очікування в обох випадках приблизно є однаковими. Проте реалізації помітно відрізняються характером зміни в часі.

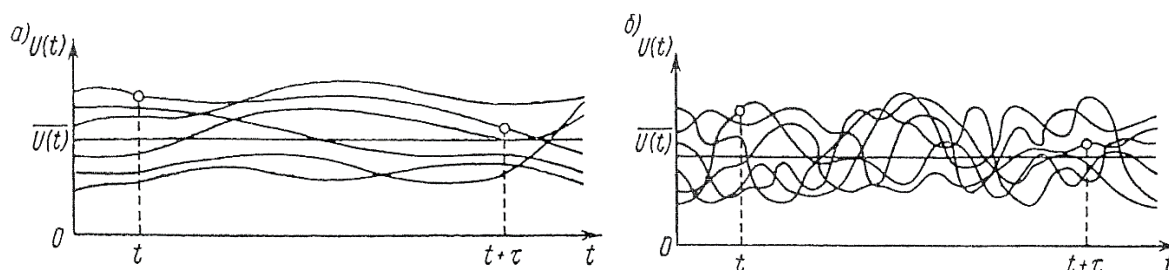


Рис. 4.7. Реалізація випадкових величин з різним ступенем кореляції: а - високим; б – малим

Функція кореляції $R(\tau)$ визначається як середнє значення добутку відхилень двох випадкових величин $\Delta U(t) = U(t) - \overline{U(t)}$ та $\Delta U(t + \tau) = U(t + \tau) - \overline{U(t + \tau)}$:

$$R(\tau) = \overline{\Delta U(t)\Delta U(t + \tau)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \Delta U(t) \cdot \Delta U(t + \tau) dt. \quad (4.21)$$

Отже, функція кореляції при стаціонарному процесі залежить тільки від величини інтервалу τ . При $\tau = 0$ величини $U(t)$ і $U(t + \tau)$ збігаються і $R(0) = \overline{\Delta U^2(t)} = D[U(t)]$. Тобто найбільше значення функції кореляції дорівнює дисперсії випадкової величини. Зі зростанням величини τ зв'язок між значеннями $U(t)$ і $U(t + \tau)$ зменшується, але не обов'язково монотонно. Типовий вид функції кореляції показано на рис. 4.8.

Функцію кореляції можна скласти і для двох різних випадкових величин, шляхом усереднення добутку значень $\Delta U(t)$ та $\Delta I(t + \tau)$. В такому випадку вона називається функцією *взаємної кореляції*.

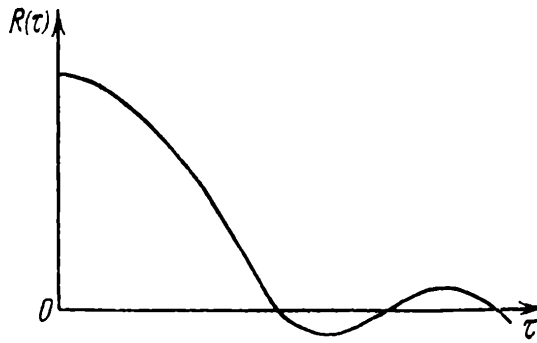


Рис. 4.8. Типовий вид функції кореляції

Отже дисперсія випадкового процесу може бути обчислена двома способами:

1. На основі відомої одновимірної щільності ймовірності

$$D_x(t) = M \left\{ [X(t) - m_x(t)]^2 \right\} = \int_{-\infty}^{\infty} x^2 p(x,t) dx - m_x^2(t).$$

2. За відомою кореляційною функцією

$$R_x(t,t) = K_x(t,t) - m_x^2(t) = D_x(t).$$

Достатньою умовою ергодичності випадкового процесу, стаціонарного в широкому сенсі, є наближення до нуля його кореляційної функції із зростанням часового зсуву τ :

$$\lim_{\tau \rightarrow \infty} R_x(\tau) = 0$$

Реальні процеси у різних системах моделюються декількома типами законів розподілу випадкових величин: рівної ймовірності, закону Сімпсона, Релея, Райса, хі-квадрат розподіл, нормальний розподіл тощо. Розглянемо ймовірнісні характеристики найбільш поширених в приладобудуванні *рівномірного і нормального* законів розподілу випадкових величин.

Для *рівномірного розподілу* щільність ймовірності є константою на деякому інтервалі $[a,b]$ (рис. 4.9, а). Ця константа згідно з умовою нормування дорівнює $1/(b-a)$.

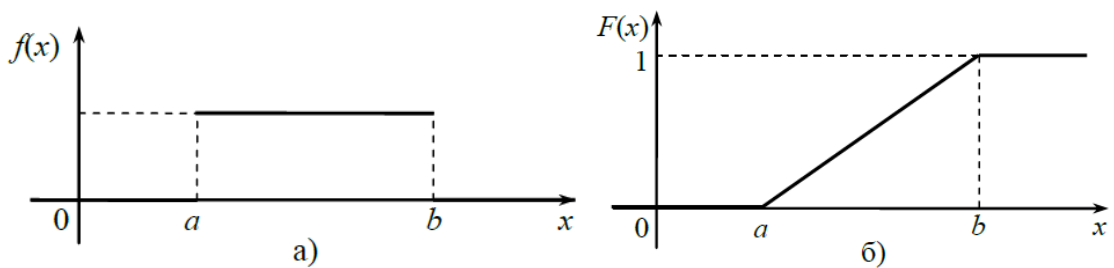


Рис. 4.9. Рівномірний закон розподілу:
 а - щільність ймовірності; б - функція розподілу

Функція розподілу на інтервалі $[a, b]$ лінійно зростає від 0 до 1 (рис. 4.9,б):

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases}$$

Математичне очікування цілком очікувано відповідає середині інтервалу можливих значень випадкової величини

$$m_x = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

Загалом, якщо функція щільності ймовірності є симетричною, тоді значення математичного очікування завжди збігається з центром симетрії.

Для розрахунку дисперсії спочатку визначимо середній квадрат випадкової величини

$$M\{X^2\} = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}.$$

За формулою (4.7) розрахуємо дисперсію

$$D_x = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Отже дисперсія рівномірно розподіленої випадкової величини дорівнює одній дванадцятій квадрата ширини інтервалу розподілу $[a, b]$. Середнє квадратичне відхилення відповідно до (4.8) є пропорційним цій ширині

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}.$$

У практиці точних вимірювань найчастіше мають справу з нормальним

розподілом результатів. Такий закон діє, коли на процес впливає одночасно велика кількість незалежних та однорідних за ступенем свого впливу випадкових факторів. При цьому вплив кожного окремого фактору є набагато меншим, ніж вплив всієї сукупності.

Для цього розподілу функція щільності розподілу та інтегральна функція:

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}, F_H(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} dx, \quad (4.22)$$

де σ_x і m_x - СКВ та математичне очікування випадкової величини x .

Особливості нормального розподілу:

1. Крива щільності розподілу є симетричною щодо ординати, що проходить через точку m_x (рис. 4.10);
2. Крива має один максимум при $x = m_x$;
3. При $|x| \rightarrow \infty$ гілки кривої асимптотично наближаються до осі абсцис;

$$\sigma_{1x} < \sigma_{2x}$$

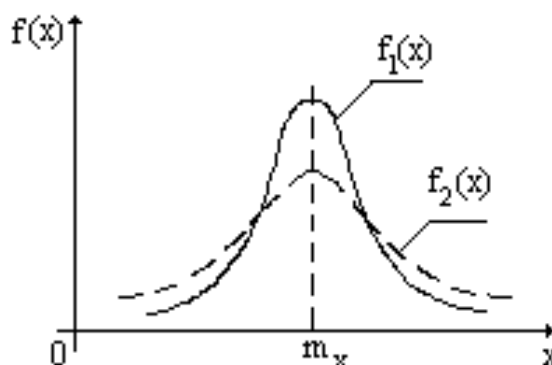


Рис. 4.10. Крива щільності нормального розподілу

$$4. \int_{\sigma_x}^{+\sigma_x} f_H(x) dx = 0,683$$

$$\int_{-2\sigma_x}^{+2\sigma_x} f_H(x) dx = 0,954$$

$$\int_{-3\sigma_x}^{+3\sigma_x} f_H(x) dx = 0,997$$

$$\int_{-\infty}^{+\infty} f_H(x) dx = 1.$$

Тобто похибка вимірювання Δ величини x в окремих вимірюваннях з ймовірністю 0,997 є обмеженою величиною $\pm 3\sigma_x$. Це значення похибки

приймається як максимальне. Вимірювання, похибка яких за абсолютним значенням перевищує $3\sigma_x$, відкидаються як явні викиди.

У разі, коли центр групування збігається з початком координат (такі величини називаються центрованими), тобто $m_x = 0$, вираз закону нормального розподілу запишеться у простішому вигляді:

$$f_H(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_x^2}}.$$

В цьому випадку ймовірність того, що величина x_i знаходиться в межах від a до b , дорівнює

$$P(a < x_i < b) = \int_a^b f(x) dx = \frac{1}{\sigma_x \sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2\sigma_x^2}} dx .$$

Вираз $e^{-\frac{x^2}{2\sigma_x^2}}$ не інтегрується в елементарних функціях. Для його визначення складено таблиці функції $\Phi(z)$, яка називається функцією Лапласа

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{z^2}{2}} dz .$$

Інтеграл $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$ називається інтегралом ймовірності.

4.3. Спектральні характеристики випадкових процесів

Кожна окрема реалізація випадкового процесу є детермінованою функцією, до якої можна застосувати перетворення Фур'є. При цьому різні реалізації будуть, природно, мати різні спектри. Тобто знаходження спектру Фур'є випадкового процесу безпосередньо неможливе, оскільки реалізації цього процесу є різними.

Отже потрібно використовувати статистично усереднені характеристики випадкових процесів. Знайдемо середнє значення спектральної щільності випадкового процесу (горизонтальною рисою тут і далі позначається операція статистичного усереднювання по ансамблю реалізацій)

$$\begin{aligned} \overline{S_x(\omega)} &= \overline{\int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt} = \int_{-\infty}^{\infty} \overline{x(t)} \exp(-j\omega t) dt = \\ &= \int_{-\infty}^{\infty} m_x(t) \exp(-j\omega t) dt. \end{aligned}$$

Як видно, усереднена спектральна щільність випадкового процесу є

спектром його детермінованої складової (математичного очікування). Для центрованих процесів $m_x(t) = 0$ і $\overline{S_x(\omega)} = 0$. Таким чином, усереднене значення спектральної щільності не несе ніякої інформації про флуктуаційну, тобто власне випадкову складову процесу. Це свідчить про випадковість і незалежність фаз спектральних складових в різних реалізаціях випадкового процесу.

Можна, проте, розглянути спектральну щільність *потужності* випадкового процесу, оскільки потужність не залежить від співвідношення фаз спектральних складових. Розглянемо центрований випадковий процес і виділимо з його ансамблю яку-небудь реалізацію $x(t)$ на обмеженому інтервалі часу $[-T/2, T/2]$. Застосувавши до цієї реалізації перетворення Фур'є, знайдемо її спектральну щільність $\dot{X}_T(\omega)$. Енергію E_T даного відрізка реалізації згідно з рівністю Парсеваля можна обчислити як

$$E_T = \int_{-T/2}^{T/2} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\dot{X}_T(\omega)|^2 d\omega.$$

Розділивши цю енергію на T , отримаємо *середню потужність* P_T реалізації на цьому часовому інтервалі

$$P_T = \langle x^2(t) \rangle_{|t| \leq T/2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\dot{X}_T(\omega)|^2}{T} d\omega. \quad (4.23)$$

Усереднивши цю потужність по ансамблю реалізацій при одночасному збільшенні тривалості проміжку часу ($T \rightarrow \infty$) отримаємо спектральну щільність середньої потужності

$$M[\langle x^2(t) \rangle] = \frac{1}{2\pi} \int_{-\infty}^{\infty} M\left[\lim_{T \rightarrow \infty} \frac{|\dot{X}_T(\omega)|^2}{T}\right] d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega) d\omega, \quad (4.24)$$

де функція $W(\omega) = M\left[\lim_{T \rightarrow \infty} \frac{|\dot{X}_T(\omega)|^2}{T}\right]$ є *спектральною щільністю середньої потужності* розглянутої реалізації (power spectral density, PSD). Застосовуються також терміни «*спектральна щільність потужності*» або «*спектр потужності*».

У випадку ергодичного процесу функція $W(\omega)$ може бути знайдена без усереднювання по ансамблю реалізацій шляхом усереднювання за часом однієї реалізації, яка в цьому випадку характеризуватиме весь процес в цілому. Середня потужність будь-якої його реалізації дорівнює дисперсії процесу

$$D_x = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega) d\omega.$$

Функція $W(\omega)$ є дійсною, вона не містить інформації про фази спектральних складових і не дозволяє відновити окремі реалізації випадкового процесу. Крім того, з визначення спектральної щільності очевидно, що $W(\omega)$ є невід'ємною і парною функцією частоти.

Тут не наводяться приклади розрахунку спектру випадкового процесу згідно з його визначенням, оскільки такого роду розрахунок не часто виконуються на практиці. Як правило, обчислення спектру випадкового процесу виконується на основі його кореляційної функції за допомогою *теорему Вінера-Хінчіна*, мова про яку піде далі.

Як розподіл спектральної щільності потужності, так і вид кореляційної функції пов'язані зі швидкістю зміни випадкового процесу в часі. Знайдемо зв'язок між цими двома характеристиками. Як відомо, кореляційна функція детермінованого сигналу пов'язана перетворенням Фур'є з його енергетичним спектром. Застосуємо цю властивість до відрізка реалізації випадкового процесу тривалістю T

$$\int_{-T/2}^{T/2} x(t)x(t-\tau)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\dot{X}_T(\omega)|^2 \exp(j\omega\tau) d\omega.$$

Розділимо обидві частини цієї рівності на T і спрямуємо T до нескінченності

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{|\dot{X}_T(\omega)|^2}{T} \exp(j\omega\tau) d\omega. \quad (4.25)$$

Якщо вважати даний процес ергодичним, то в лівій частині останньої рівності стоїть кореляційна функція процесу, отримана шляхом усереднювання за часом. У правій частині під інтегралом міститься вираз $W(\omega)$ для спектральної щільності потужності випадкового процесу. З урахуванням цього маємо

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega) \exp(j\omega\tau) d\omega. \quad (4.26)$$

У разі неергодичного процесу до обох частин рівності (4.25) необхідно додатково застосувати усереднювання по ансамблю реалізацій, що приведе до того ж самого результату.

Таким чином, кореляційна функція центрованого стаціонарного випадкового процесу і його спектральна щільність потужності пов'язані

перетвореннями Фур'є.

Тобто при відомій кореляційній функції спектральна щільність потужності може бути визначена як

$$W(\omega) = \int_{-\infty}^{\infty} R(\tau) \exp(-j\omega\tau) d\tau. \quad (4.27)$$

Співвідношення (4.26) і (4.27) складають зміст *теорему Вінера-Хінчіна*, доведеної в тридцять років минулого століття незалежно один від одного американським і радянським математиками Н. Вінером і О. Я. Хінчіним.

Оскільки як $W(\omega)$, так і $R(\tau)$ є парними дійсними функціями, для випадкових процесів з нульовим середнім можна відмовитись від комплексної форми запису перетворення Фур'є і перейти до напівнескінчених границь інтегрування:

$$R(\tau) = \frac{1}{\pi} \int_0^{\infty} W(\omega) \cos(\omega\tau) d\omega, \quad W(\omega) = 2 \int_0^{\infty} R(\tau) \cos(\omega\tau) d\tau. \quad (4.28)$$

При $\tau = 0$ маємо

$$R(0) = \frac{1}{\pi} \int_0^{\infty} W(\omega) d\omega = \sigma^2.$$

Дисперсія стаціонарного випадкового процесу, таким чином, є сумою внесків від усіх ділянок частотної вісі. На відміну від детермінованих імпульсних сигналів $W(\omega)$ характеризує питому міру потужності, а не частку енергії, яка припадає на одиничну смугу частот.

Для зручності розрахунків також вводять односторонній спектр потужності $F(\omega)$ випадкового процесу

$$F(\omega) = \begin{cases} 0, & \omega < 0, \\ W(\omega)/\pi, & \omega \geq 0. \end{cases} \quad (4.29)$$

Представлення $F(\omega)$ з допомогою (4.29) дозволяє знаходити дисперсію стаціонарного випадкового процесу інтегруванням тільки по фізично існуючим додатним частотам

$$\sigma^2 = R(0) = \int_0^{\infty} F(\omega) d\omega. \quad (4.30)$$

В техніці також широко застосовується односторонній спектр потужності $N(f)$, який характеризує середню потужність випадкового процесу на кожному з інтервалів частот в 1 Гц

$$N(f) = \begin{cases} 0, & f < 0, \\ 2W(2\pi f), & f \geq 0. \end{cases} \quad (4.31)$$

В цьому випадку

$$\sigma^2 = \int_0^\infty N(f)df. \quad (4.32)$$

Випадкові процеси, які розглядаються в теорії сигналів, часто мають функцію кореляції, що наближається до нуля із збільшенням часового зсуву τ (нагадаємо, що це є достатньою умовою ергодичності процесу) (рис. 4.11). Чим швидше убуває функція $R(\tau)$, тим слабкішим є статистичний зв'язок між миттєвими значеннями випадкового сигналу в два не співпадаючі моменти часу. Числовою характеристикою для оцінки «швидкості зміни» реалізацій випадкового процесу слугує *інтервал кореляції* τ_k , який визначається за формулою

$$\tau_k = \frac{1}{R(0)} \int_0^\infty |R(\tau)| d\tau = \int_0^\infty |r(\tau)| d\tau,$$

де $r(\tau)$ – нормована функція кореляції.

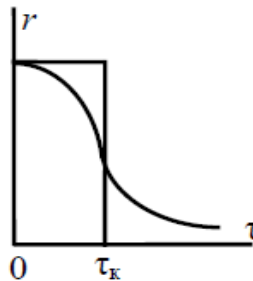


Рис. 4.11. До визначення інтервалу кореляції: площі обох фігур однакові

Якщо відома інформація про поведінку якої-небудь реалізації випадкового процесу «у минулому», то можливий ймовірнісний прогноз випадкового процесу на час порядку τ_k . Прогнозування на інтервал часу, який суттєво перевищує τ_k , виявляється безрезультатним, оскільки миттєві значення стають практично некорельованими.

З іншого боку, досліджуваний випадковий процес характеризується спектром щільності потужності $W(\omega)$, що має максимальне значення W_{\max} . Подібно до методики знаходження інтервалу кореляції (рис. 34) замінімо подумки цей випадковий процес іншим, у якого спектральна щільність потужності постійна і дорівнює W_{\max} в межах деякої смуги частот, яка обрана з умови рівності дисперсій (тобто середніх потужностей) обох процесів. Ширина цієї смуги частот називається *ефективною шириною спектру* випадкового процесу і має вигляд:

$$W_{max}\Delta\omega_{ef} = \int_{-\infty}^{\infty} W(\omega)d\omega, \Delta\omega_{ef} = \frac{1}{W_{max}} \int_{-\infty}^{\infty} W(\omega)d\omega. \quad (4.33)$$

Білий шум.

В радіовимірюваннях та радіотехніці «білим шумом» називають стаціонарний випадковий процес, спектральна щільність потужності якого є постійною на всіх частотах

$$W(\omega) = W_0 = const.$$

Згідно з теоремою Вінера-Хінчіна кореляційна функція білого шуму є дельта-функцією

$$R(\tau) = \frac{W_0}{2\pi} \int_{-\infty}^{\infty} \exp(j\omega\tau) d\omega = W_0\delta(\tau),$$

тобто дорівнює нулю усюди, окрім точки $\tau = 0$.

Дисперсія білого шуму є нескінченно великою. Значення білого шуму у різні моменти часу є некорельованими, оскільки, яким би не був малим інтервал τ , сигнал за цей час може змінитися на будь-яку величину.

Білий шум є абстрактною математичною моделлю і фізично існувати не може. Це пояснюється передусім нескінченністю його дисперсії (тобто середньої потужності). Проте в тих випадках, коли смуга пропускання досліджуваної системи істотно вужча за ефективну ширину спектру шуму, який на неї впливає, можна для спрощення аналізу приблизно замінити реальний випадковий процес білим шумом.

Питання для самоперевірки

1. Дайте визначення випадкового сигналу.
2. Класифікація випадкових сигналів.
3. Ознаки стаціонарних випадкових процесів.
4. Стаціонарні та ергодичні випадкові процеси.
5. Що таке закон розподілу випадкового процесу?
6. Що таке диференціальний закон розподілу?
7. Що таке початкові моменти в теорії ймовірностей?
8. Що таке центральні моменти в теорії ймовірностей?
9. Математичне очікування та середнє значення випадкового процесу.
10. Сутність теореми Вінера-Хінчіна.
11. Рівність Парсеваля.

12. Які характеристики використовують для оцінки початкових та центральних моментів розподілу?
13. Як пов'язана кореляційна функція з дисперсією випадкового процесу?
14. Зв'язок кореляційної функції з спектральною щільністю.
15. Які випадкові процеси називають ергодичними?
16. Які властивості має нормальний розподіл випадкової величини?

РОЗДІЛ 5. СИСТЕМИ ОБРОБКИ СИГНАЛІВ

Сучасний розвиток цифрової елементної бази та зручність використання цифрових засобів вимірювання призводять до масового їх втілення в різних галузях вимірювань. В той же час багато базових понять і методів, які використовуються при цифровій обробці дискретних сигналів, прийшли з певними особливостями від аналогових систем.

Аналогові засоби вимірювання нині також застосовуються, а в цифрових засобах можуть використовуватись окремі аналогові елементи. Тому уявлення про характеристики і способи описання аналогових систем також потрібне, як і для дискретних.

Це дозволить глибше зрозуміти й теорію дискретних систем, тому що багато методів аналізу аналогових і дискретних систем знаходяться в тісній спорідненості.

5.1. Аналогові системи обробки сигналів

Системи, що застосовуються для перетворення сигналів, мають різноманітні фізичні характеристики і можуть класифікуватися за різними ознаками. Найважливішою класифікаційною ознакою є лінійність або нелінійність системи.

Лінійними називаються системи, для яких виконується принцип суперпозиції: реакція на суму сигналів дорівнює сумі реакцій на ці сигнали, подані на вхід окремо. Так, якщо зв'язок між сигналами на вході і виході системи задається оператором T

$$\mathbf{U}_{\text{вих}}(t) = T[\mathbf{U}_{\text{вх}}(t)],$$

де $\mathbf{U}_{\text{вих}}(t) = (s_{\text{вих}1}(t)s_{\text{вих}2}(t) \dots s_{\text{вих}n}(t))$ – вихідний n -мірний векторний сигнал; $\mathbf{U}_{\text{вх}}(t) = (s_{\text{вх}1}(t)s_{\text{вх}2}(t) \dots s_{\text{вх}m}(t))$ – вхідний m -мірний векторний сигнал,

то для лінійної системи справедливим є:

$$T[\mathbf{U}_{\text{вх}1}(t) + \mathbf{U}_{\text{вх}2}(t)] = T[\mathbf{U}_{\text{вх}1}(t)] + T[\mathbf{U}_{\text{вх}2}(t)];$$

$$T(a\mathbf{U}_{\text{вх}}(t)) = aT(\mathbf{U}_{\text{вх}}(t)).$$

Системи, для яких принцип суперпозиції не виконується, називаються *нелінійними*.

Наступним критерієм класифікації систем є постійність або змінність їх характеристик в часі. Якщо довільна затримка вхідного сигналу призводить лише до такої ж затримки вихідного сигналу, не змінюючи його форми, система називається *стаціонарною*.

Тобто якщо дискретна система визначається виразом $U_{\text{вих}}(t) = T[U_{\text{вх}}(t)]$, то вона називається стаціонарною, якщо для будь-якої вхідної послідовності $U_{\text{вх}}(t)$ і довільного цілого числа t_0 виконується співвідношення $U_{\text{вих}}(t \pm t_0) = T[U_{\text{вх}}(t \pm t_0)]$.

Інакше система називається *нестаціонарною*, параметричною або системою із змінними параметрами.

Два вказані способи класифікації ділять системи на чотири класи: лінійні стаціонарні, лінійні не стаціонарні, нелінійні стаціонарні та нелінійні нестаціонарні.

Далі будемо розглядати лінійні стаціонарні системи.

Систему називають *детермінованою*, якщо члени вихідної послідовності з номером n_0 залежать лише від тих членів вхідної послідовності, номера яких не перевищують n_0 . Ця умова визначає той факт, що якщо $x_1[n] = x_2[n]$ при $n \leq n_0$, то $y_1[n] = y_2[n]$ при $n \leq n_0$.

Права різницєва система, визначається за правилом:

$$y[n] = x[n + 1] - x[n]. \quad (5.1)$$

Вона не є детермінованою, оскільки кожний із членів вихідної послідовності з номером n розраховується як за $x[n]$ так і за $x[n+1]$. Порушення детермінованості легко помітити, якщо взяти в якості вхідних послідовностей $x_1[n]=\delta[n-1]$ та $x_2[n]=0$. Відповідні відгуки - $y_1[n]=\delta[n]-\delta[n-1]$ та $y_2[n]=0$. Відмітимо, що при $x_1[n]=x_2[n]$ при $n \leq 0$. Тому відповідно до визначення детермінованості, повинна виконуватись рівність $y_1[n]=y_2[n]$ для всіх $n \leq 0$, яке порушується при $n=0$.

Ліва різницєва система визначається як

$$y[n] = x[n] - x[n - 1]. \quad (5.2)$$

Кожний член вихідної послідовності такої системи залежить від члена з тим самим номером вхідної послідовності та одного попереднього. Тому вона є детермінованою.

Стійкість. Говорять, що система є стійкою, якщо і лише якщо її реакція на довільний обмежений за амплітудою сигнал є обмеженою. Послідовність $x[n]$ називається обмеженою, якщо знайдеться таке додатне число B_x , що для будь-якого n

$$|x[n]| \leq B_x < \infty.$$

Таким чином, для стійкої системи кожна обмежена послідовність має у відповідності таку ж додатну константу B_y , що для будь-якого n

$$|y[n]| \leq B_y < \infty.$$

Дані властивості, це властивості самої системи, а не вхідної послідовності. Інакше кажучи, для будь-якої системи, можуть існувати пари сигнал-відгук, що забезпечують виконання даної умови. Однак, наявність таких пар ще не визначає стійкість системи. Дана система буде стійкою лише в тому випадку, коли дану властивість будуть мати всі пари.

Розглянемо різні еквівалентні способи описання лінійних систем із зосередженими параметрами та переходу від одного представлення до іншого, що необхідно для правильного їх застосування на практиці.

Для лінійних систем з постійними параметрами справедливі принципи суперпозиції і стаціонарності. Це значно спрощує аналіз проходження сигналів через такі системи. Лінійність і стаціонарність дозволяють легко знайти реакцію системи на будь-який вхідний сигнал, знаючи лише одну функцію – реакцію системи на подану на вхід δ -функцію. Ця реакція називається *імпульсною характеристикою* системи і позначається $h(t)$. Якщо деяка лінійна стаціонарна система описується оператором \mathbf{T} і для спрощення вважати вхідний і вихідний сигнали одновимірними, то $h(t) = \mathbf{T}[\delta(t)]$. Оскільки система є стаціонарною, то аналогічний вихідний сигнал матиме місце і для зсунутої на довільний час δ -функції $\delta(t - t_0)$:

$$h(t - t_0) = \mathbf{T}[\delta(t - t_0)].$$

З використанням динамічного представлення представимо сигнал на вході лінійної системи у вигляді його згортки з δ -функцією

$$s_{\text{вх}}(t) = \int_{-\infty}^{\infty} s_{\text{вх}}(t') \delta(t - t') dt'.$$

Лінійна система під дією оператора \mathbf{T} перетворює вхідний сигнал $s_{\text{вх}}(t)$

відносно змінної t , тобто усі функції, що входять в цей вираз і залежать від t . Вхідний сигнал $s_{\text{вх}}(t)$ при цьому перетворюється на вихідний сигнал $s_{\text{вих}}(t)$ а δ -функція $\delta(t-t')$ – в імпульсну характеристику $h(t-t')$.

Функція $s_{\text{вх}}(t')$ від t не залежить і тому залишається без змін. В результаті отримуємо формулу, яка показує, що сигнал на виході лінійної системи з постійними параметрами дорівнює згортці вхідного сигналу й імпульсної характеристики системи

$$s_{\text{вих}}(t) = \mathbf{T}[s_{\text{вх}}(t)] = \mathbf{T}\left[\int_{-\infty}^{\infty} s_{\text{вх}}(t')\delta(t-t')dt'\right] = \int_{-\infty}^{\infty} s_{\text{вх}}(t')\mathbf{T}[\delta(t-t')]dt' = \int_{-\infty}^{\infty} s_{\text{вх}}(t')h(t-t')dt'. \quad (5.3)$$

Ця формула є фундаментальною в теорії лінійних систем і має назву *інтеграла Дюамеля*. Вона також може бути подана у вигляді

$$s_{\text{вих}}(t) = \int_{-\infty}^{\infty} s_{\text{вх}}(t-t')h(t')dt'.$$

Зазначимо, що, якщо вхідний і вихідний сигнали системи мають однакову розмірність, імпульсна характеристика, як і δ -функція часу, має розмірність частоти.

Якщо $h(t)$ відома заздалегіть, то подальші етапи знаходження вихідного сигналу системи є цілком формалізованими.

Для випадку багатовимірного сигналу вводяться *парціальні* імпульсні характеристики $h_{ij}(t)$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$, кожна з яких відображає сигнал на i -тому виході при подачі на j -й вхід δ -функції. Сукупність $h_{ij}(t)$ утворюють матрицю імпульсних характеристик

$$\mathbf{h}(t) = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1m} \\ h_{21} & h_{22} & \dots & h_{2m} \\ \cdot & \cdot & \dots & \cdot \\ h_{n1} & h_{n2} & \dots & h_{nm} \end{pmatrix}.$$

Формула інтеграла Дюамеля в багатовимірному випадку має вигляд

$$\mathbf{U}_{\text{вих}}(t) = \int_{-\infty}^{\infty} \mathbf{h}(t-t')\mathbf{U}_{\text{вх}}(t')dt', \text{ де } \mathbf{U}_{\text{вх}}(t) - m\text{-вимірний вектор, } \mathbf{U}_{\text{вих}}(t) - n\text{-}$$

вимірний вектор.

Формування вихідного сигналу можна пояснити таким чином.

Нескінченно мала «ділянка» вхідного сигналу $s_{\text{вх}}(t')$ шириною dt' породжує на виході відгук, що є імпульсною характеристикою, помноженою на $s_{\text{вх}}(t')dt'$ і затриманою за часом на t' , тобто є $s_{\text{вх}}(t')h(t-t')dt'$ (рис. 5.1) [4]. Щоб набути значення вихідного сигналу у момент часу t , треба скласти внески від усіх цих нескінченно малих «ділянок», тобто виконати інтегрування по t' , що і дає приведену вище формулу згортки (5.1).

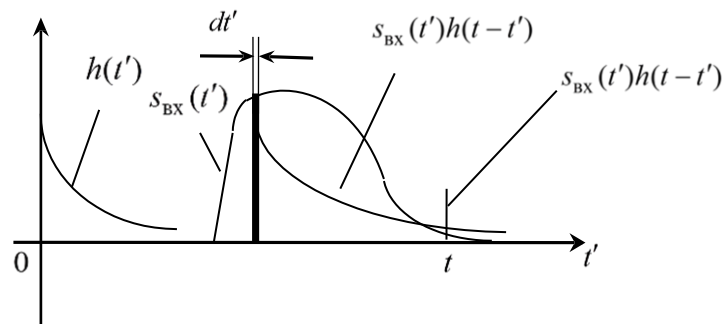


Рис. 5.1. Формування вихідної реакції лінійної системи

Умовою фізичної реалізованості системи є важливий принцип: сигнал на виході системи, який відповідає вхідному впливові, не може з'явитися до моменту появи імпульсу на вході. Звідси маємо просте обмеження на вид можливих імпульсних характеристик: $h(t)=0$ при $t<0$. Для системи, що фізично реалізується, верхня границя в інтегралі Дюамеля може бути замінена на поточний час

$$s_{\text{вих}}(t) = \int_{-\infty}^t s_{\text{вх}}(t') h(t-t') dt'.$$

Ця формула має ясний фізичний зміст: лінійна стаціонарна система при обробці вхідного сигналу виконує операцію вагового додавання всіх його миттєвих значень, які існували «в минулому» при $-\infty < t' < t$. Роль вагової функції виконує при цьому імпульсна характеристика системи. Принципово важливим є те, що система, яка реалізується фізично, ні за яких обставин не здатна оперувати з «майбутніми» значеннями вхідного сигналу. Окрім цього, вона має бути стійкою. Тобто, її імпульсна характеристика повинна задовольняти умові абсолютної інтегрованості

$$\int_{-\infty}^{\infty} |h(t)| dt < \infty.$$

Перехідною характеристикою називають реакцію системи на подану

на вхід функцію одиничного стрибка (функцію Хевісайда). Позначається перехідна характеристика як $g(t)=\mathbf{T}[\sigma(t)]$. Перехідна характеристика стаціонарної системи є інваріантною відносно часового зсуву:

$$g(t-t_0)=\mathbf{T}[\sigma(t-t_0)].$$

Перехідна характеристика системи, що фізично реалізується, відрізняється від нуля лише при $t \geq 0$, як і для імпульсної характеристики $g(t)=0$ при $t < 0$.

Оскільки δ -функція є похідною від одиничного стрибка, імпульсна і перехідна характеристики пов'язані одна з одною операціями диференціювання й інтегрування:

$$h(t)=\frac{dg(t)}{dt}; \quad g(t)=\int_{-\infty}^t h(t') dt'. \quad (5.4)$$

Скориставшись динамічним представленням довільного сигналу із застосуванням функцій Хевісайда

$$s(t)=s_0\sigma(t)+\int_0^{\infty} \frac{ds}{dt'} \sigma(t-t') dt' \quad \text{— (розділ 1), отримаємо форму інтеграла}$$

Дюамеля у вигляді

$$s_{\text{вих}}(t)=\mathbf{T}[s_{\text{вх}}(t)]=s_{\text{вх}}(0)g(t)+\int_0^t \frac{s_{\text{вх}}(t')}{dt'} g(t-t') dt'. \quad (5.5)$$

Рівноцінною є й інша форма представлення сигналу на виході лінійної стаціонарної системи інтегралом Дюамеля такого виду

$$s_{\text{вих}}(t)=s_{\text{вх}}(t)g(0)+\int_0^t \frac{g(t')}{dt'} s_{\text{вх}}(t-t') dt'.$$

Оскільки сигнал на виході лінійної системи, як було показано вище, є згортокою вхідного сигналу та імпульсної характеристики, то перетворення Фур'є від згортки сигналів дає добуток їх спектрів. Спектр сигналу на виході лінійної системи в результаті виглядає як

$$\dot{S}_{\text{вих}}(\omega)=\dot{S}_{\text{вх}}(\omega)\dot{K}(\omega). \quad (5.6)$$

де $\dot{K}(\omega)$ – перетворення Фур'є імпульсної характеристики системи

$$\dot{K}(\omega)=\int_{-\infty}^{\infty} h(t) e^{-j\omega t} dt. \quad (5.7)$$

Ця функція називається *комплексним (частотним) коефіцієнтом передачі* системи, а її модуль $|\dot{K}(\omega)|$ і фаза $\varphi_K(\omega)=\arg[\dot{K}(\omega)]$ – відповідно

амплітудно-частотною (АЧХ) і фазочастотною (ФЧХ) характеристиками системи.

Значення $\dot{K}(\omega)$ показує, як змінюється при проходженні через систему комплексна амплітуда синусоїди з частотою ω . Відповідно АЧХ показує у скільки разів зміниться амплітуда синусоїди, а ФЧХ – який буде отриманий нею фазовий зсув

$$\dot{K}(\omega) = |\dot{K}(\omega)| \exp[j\varphi_K(\omega)].$$

Оскільки $h(t)$ – функція дійсна, то з урахуванням властивостей перетворення Фур'є $\dot{K}(\omega) = \dot{K}^*(-\omega)$. Тому АЧХ є парною, а ФЧХ – непарною функціями частоти.

Потужність гармонічного сигналу пропорційна квадрату його амплітуди і не залежить від його фази. Тому *коефіцієнт передачі за потужністю* дорівнює квадрату модуля комплексного коефіцієнта передачі, тобто квадрату АЧХ

$$\dot{K}_{\text{пот}}(\omega) = \dot{K}(\omega)\dot{K}^*(\omega) = |\dot{K}(\omega)|^2. \quad (5.8)$$

При перетворенні сигналу лінійною системою розрізняють два види затримки. Фазова затримка (phase delay) на частоті ω – це затримка гармонійного коливання з частотою ω , що проходить через систему. Значення фазової затримки дорівнює фазовому зсуву, що вноситься системою, діленому на частоту гармонійного коливання, із від'ємним знаком: $\tau_{\text{ф}}(\omega) = -\varphi_K(\omega)/\omega$.

Групова затримка (group delay) на частоті ω – це затримка обвідної вузько-смугового сигналу з середньою частотою ω . Групова затримка дорівнює похідній від ФЧХ системи із від'ємним знаком: $\tau_{\text{гр}}(\omega) = -d\varphi_K(\omega)/d\omega$.

Взаємний спектр вихідного і вхідного сигналів лінійної системи можна знайти, виходячи з визначення взаємного спектра $\dot{S}_{12}(\omega) = \dot{S}_1(\omega)\dot{S}_2^*(\omega)$:

$$\dot{S}_{\text{вих_вих}}(\omega) = \dot{S}_{\text{вих}}(\omega)\dot{S}_{\text{вих}}^*(\omega) = \dot{S}_{\text{вих}}(\omega)\dot{K}(\omega)\dot{S}_{\text{вих}}^*(\omega) = |\dot{S}_{\text{вих}}(\omega)|^2 \dot{K}(\omega).$$

Звідси витікає, що комплексний коефіцієнт передачі системи дорівнює відношенню взаємного спектру вихідного і вхідного сигналів до енергетичного спектру вхідного сигналу:

$$\dot{K}(\omega) = \frac{\dot{S}_{\text{вих_вих}}(\omega)}{|\dot{S}_{\text{вих}}(\omega)|^2}. \quad (5.9)$$

Взаємна кореляція між входом і виходом може бути отримана застосуванням оберненого перетворення Фур'є до формули для взаємного спектру. Вираз для ВКФ сигналів на виході і вході з урахуванням зв'язку між кореляційною функцією і спектром сигналу має вигляд

$$B_{\text{вих_вих}}(\tau) = \int_{-\infty}^{\infty} B_{\text{вих}}(\tau') h(\tau - \tau') d\tau'.$$

Отже, ВКФ сигналів на вході і виході лінійної системи є згортокою АКФ вхідного сигналу з імпульсною характеристикою системи.

Зв'язок між сигналами на вході і виході лінійного ланцюга із зосередженими параметрами може бути поданий диференціальним рівнянням виду

$$\begin{aligned} a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + a_{n-2} \frac{d^{n-2} y}{dt^{n-2}} + \dots + a_1 \frac{dy}{dt} + a_0 y(t) = \\ b_m \frac{d^m x}{dt^m} + b_{m-1} \frac{d^{m-1} x}{dt^{m-1}} + a_{m-2} \frac{d^{m-2} x}{dt^{m-2}} + \dots + b_1 \frac{dx}{dt} + b_0 x(t), \end{aligned} \quad (5.10)$$

де $x(t)$ – вхідний сигнал, $y(t)$ – вихідний сигнал, a_i і b_i – постійні коефіцієнти. Таким чином, ланцюг описується наборами коефіцієнтів $\{a_i\}$ і $\{b_i\}$.

Повинна виконуватись нерівність $m \leq n$, тобто максимальний порядок похідної сигналу на вході не може перевищувати максимального порядку похідної сигналу на виході. Це пов'язано з неможливістю реалізації операції «чистого» диференціювання аналоговим ланцюгом. Значення n називається *порядком динамічної системи*.

Якщо задати конкретний вид вхідного сигналу $x(t)$, вийде лінійне неоднорідне диференціальне рівняння з постійними коефіцієнтами. Рішення цього ДУ дає вихідний сигнал $y(t)$.

Застосувавши до обох частин диференціального рівняння (5.10) перетворення Лапласа, отримаємо вираз для операторного коефіцієнта передачі, або *передавальної функції динамічної системи* (transfer function)

$$H(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + b_{m-2} s^{m-2} + \dots + b_1 s + b_0}{a_n s^n + a_{n-1} s^{n-1} + a_{n-2} s^{n-2} + \dots + a_1 s + a_0} \quad (5.11)$$

Як видно з (5.11), $H(s) = \frac{Y(s)}{X(s)}$ є відношенням перетворень Лапласа сигналу на виході $Y(s)$ та на вході $X(s)$ системи.

5.2. Аналогові фільтри

В аналоговій обробці сигналів під фільтрами розуміють пристрої для виділення бажаних компонентів спектру електричного сигналу та подавлення не бажаних компонентів.

За формою графіку модуля комплексної частотної характеристики (за видом АЧХ) фільтри ділять на чотири основні типи:

1) фільтр нижніх частот (ФНЧ, *low-pass filter*) – пропускає частоти, менші деякої частоти зрізу ω_0 .

Такий фільтр має порівняно великий коефіцієнт передачі на частотах, які знаходяться в околі нульової частоти, і порівняно низький – на інших частотах. Тому такі фільтри зберігають незмінною величину спектральних складових вхідного сигналу на нижніх частотах; говорять, що ФНЧ пропускають низькочастотні сигнали та не пропускають високочастотні;

2) фільтр верхніх частот (ФВЧ, *high-pass filter*) – пропускає частоти, більші деякої частоти зрізу ω_0 . Такий фільтр має високий коефіцієнт передачі на частотах, які більше за ω_0 , і низький – на частотах від нуля до ω_0 . Отже вони пропускають без змін частину спектру вхідного сигналу, яка знаходиться на частотах від ω_0 до $+\infty$ і не пропускають складові спектру сигналу від 0 до ω_0 ;

3) смуговий фільтр (СФ, *band-pass filter*) – має великий коефіцієнт передачі лише в певній смузі частот між частотами ω_1 та ω_2 . Поза цим проміжком коефіцієнт передачі є малим. Отже, на вихід такого фільтра пройдуть лише ті спектральні складові сигналу, які знаходяться в смузі пропускання фільтра, і не пройдуть ті, які лежать від 0 до ω_1 та від ω_2 до $+\infty$. Даний фільтр також характеризується центральною частотою $\omega_0 = (\omega_1 + \omega_2)/2$ та шириною смуги пропускання $\Delta\omega = (\omega_2 - \omega_1)$;

4) загороджувальний фільтр (ЗФ, режекторний фільтр, *band-stop filter*) – пропускає на вихід всі частоти, крім тих, що лежать в деякому діапазоні ($\omega_1 - \omega_2$). Він має великий коефіцієнт передачі поза проміжком

частот ($\omega_1 - \omega_2$). Якщо смуга затримки такого фільтра порівняно вузька, такий загороджувальний фільтр називають іноді фільтром-пробкою (*notch-filter*).

Проміжок частот, для яких коефіцієнт передачі фільтра є великим, називається смугою пропускання (*passband*). Якщо спектральні складові вхідного сигналу фільтра потрапляють в смугу пропускання, то вони не зазнають ослаблення, проходять на вихід фільтра без змін та залишаються в спектрі вихідного сигналу.

Проміжок частот, для яких коефіцієнт передачі фільтра є малим (для ідеального фільтра він дорівнює нулю), називається смугою затримки (*stopband*). Якщо частина спектру сигналу потрапляє в смугу затримки, то ці складові зазнають значного ослаблення, і у вихідному сигналі фільтра вони присутні не будуть. В цьому випадку говорять, що ця частина спектру сигналу затримана фільтром.

В реальних фільтрах неможливо забезпечити різкий перехід між смугою пропускання та смугою затримки. Між цими двома смугами буде знаходитися частотний проміжок, на якому коефіцієнт передачі змінюється – перехідна смуга (*transition band*). АЧХ ідеальних і реальних фільтрів подано на рис. 5.2 та 5.3 відповідно.

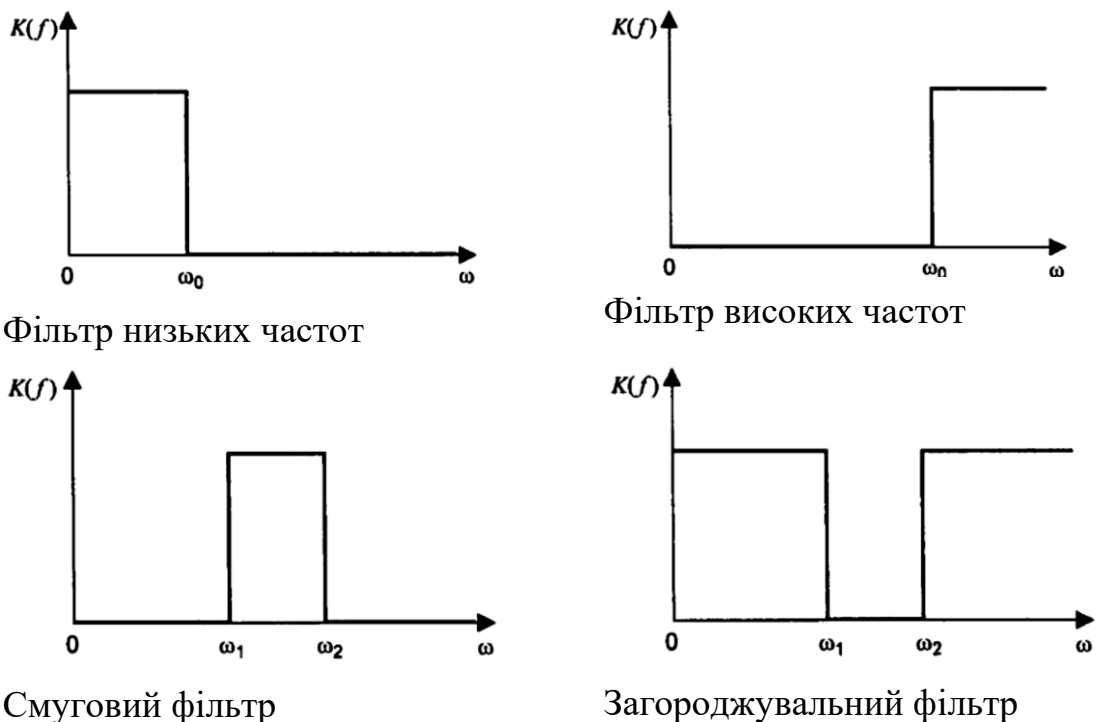


Рис. 5.2. АЧХ ідеальних аналогових фільтрів

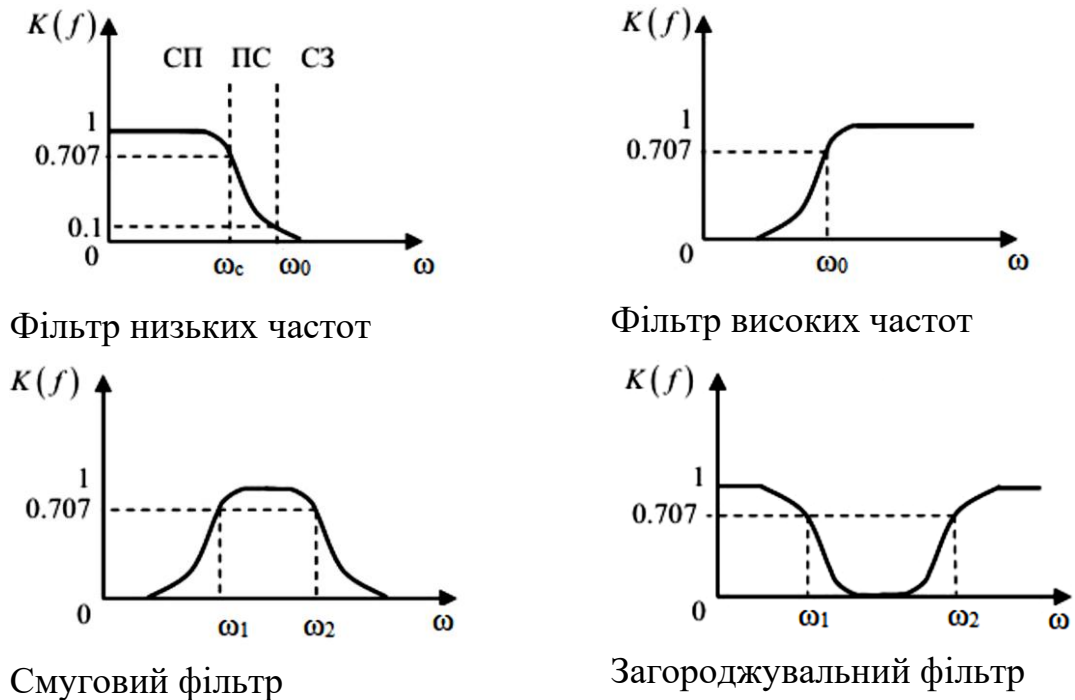


Рис. 5.3. АЧХ реальних аналогових фільтрів

Видно, що зміна коефіцієнта передачі для таких фільтрів відбувається не раптово. Смуга пропускання реальних фільтрів визначається по-іншому: це смуга, в якій фільтр має достатньо великий коефіцієнт передачі. Вважається, що це смуга частот, для яких коефіцієнт передачі зменшується не менше, ніж в $\sqrt{2}$ разів.

На основі цього визначається *частота зрізу фільтра* – це та частота, на якій коефіцієнт передачі зменшується в $\sqrt{2}$ разів відносно максимального коефіцієнта передачі в смузі пропускання. Для випадку, коли коефіцієнт передачі фільтра вимірюється в децибелах, то частота зрізу визначається на рівні -3 дБ від максимального значення.

Оскільки ідеальна форма АЧХ не може бути фізично реалізованою, в теорії аналогових фільтрів розроблено ряд методів апроксимації прямокутних АЧХ.

За функцією апроксимації АЧХ фільтри розрізняються:

- фільтри Батерворта – головною перевагою яких є відсутність пульсацій в смузі пропускання, проте вони мають повільний спад АЧХ в смузі затримки, тому реалізувати високо вибіркові фільтри складно та дорого, оскільки потрібно забезпечити високий порядок фільтру;

- фільтр Чебишева першого та другого роду – мають значно крутіший спад АЧХ, проте мають пульсації в смузі пропускання (фільтри 1 роду) або в смузі затримки (фільтри 2 роду).

- еліптичні фільтри (фільтри Кауера) – мають найвищу крутість спаду АЧХ, проте є пульсації як в смузі пропускання, так і в смузі затримки з більш низькою ефективністю придушення в ній.

Також розрізняють фільтри Бесселя (які мають гладку фазо-частотну характеристику), Лагранжа, Габора, Гауса тощо.

Окрім того розраховувачи ФНЧ, можна шляхом нескладних перетворень змінити частоту зрізу, перетворивши його у ФВЧ, ПФ або ЗФ із заданими параметрами. Тому розрахунок аналогового фільтру починається з розрахунку так званого фільтру-прототипу, що представляє собою ФНЧ з частотою зрізу, що дорівнює 1рад/с. В подальшому застосовуються функції перетворення частоти зрізу і перетворення типу фільтру.

5.3. Дискретні системи обробки сигналів

З точки зору математики система з дискретним часом визначається як перетворення, або оператор, що переводить вхідну послідовність (сигнал) $x(n)$ у вихідну послідовність $y(n)$ (відгук або реакцію системи), що можна представити як:

$$Y(n) = T\{x(n)\}. \quad (5.12)$$

Дане співвідношення – це правило за яким визначається значення реакції системи через відліки сигналу, що подається на її вхід. Необхідно зазначити, що відлік реакції системи з індексом n може залежати від усіх відліків вхідного сигналу $x(n)$.

Для описання дискретних лінійних систем (фільтрів) подібно до аналогових широко застосовуються їх імпульсні характеристики та системні функції (передавальні функції).

Імпульсною характеристикою $h(n)$ дискретного фільтра називається його реакція на одиничний імпульс $u_0(n)$ при нульових початкових умовах (подібно до аналогових систем, де вхідний вплив – δ -функція)

$$h(n) = F[u_0(n)], \text{ або } h(k, n) = h(n - k) = F[u_0(n - k)]. \quad (5.13)$$

де F – оператор перетворення дискретного фільтра.

Імпульсну характеристику фільтра $h(k, n)$ зазвичай представляють сукупністю значень $\{h(0), h(1), h(2), \dots\}$.

Зазначимо, що будь-яку дискретну послідовність $x(n)$ можна представити у вигляді лінійної комбінації одиничних відліків

$$x(n) = \sum_{k=-\infty}^{\infty} x(k) u_0(n-k). \quad (5.14)$$

Тоді сигнал на виході фільтра, виходячи з лінійності і стаціонарності такої системи, повинен представляти собою лінійну комбінацію імпульсних характеристик

$$\begin{aligned} y(n) &= F[x(n)] = F\left[\sum_{k=-\infty}^{\infty} x(k) u_0(n-k)\right] = \sum_{k=-\infty}^{\infty} x(k) F[u_0(n-k)] = \\ &= \sum_{k=-\infty}^{\infty} x(k) h(n-k) = \sum_{k=-\infty}^{\infty} x(n-k) h(k). \end{aligned} \quad (5.15)$$

Вираз (6.15) є *дискретною згорткою* вхідного сигналу та імпульсної характеристики фільтра (яку в цифровій техніці називають також оператором згортки). Якщо фільтр каузальний, тобто реалізується фізично, тоді реакція на його виході не може з'явитись раніше впливу на вході (принцип часового детермінізму). Як наслідок, імпульсна характеристика такого фільтра повинна дорівнювати нулю у відлікових точках, які передують моменту подачі вхідного імпульсу, тобто $h(n) = 0$ при $n < 0$. Тому верхню границю підсумовування в формулі (6.15) може бути замінено на n

$$y(n) = \sum_{k=0}^n x(k) h(n-k) = \sum_{k=0}^n x(n-k) h(k). \quad (5.16)$$

Застосуємо третю властивість z -перетворення до рівняння дискретної фільтрації: $y(n) = \sum_{k=-\infty}^{\infty} x(k) h(n-k)$. Оскільки вихідний сигнал дискретної системи є лінійною згорткою вхідного сигналу з імпульсною характеристикою системи, тоді

$$Y(z) = X(z) \cdot H(z), \quad (5.17)$$

де $X(z), Y(z)$ - z -перетворення відповідно вхідного і вихідного сигналів системи, а $H(z)$ - z -перетворення її імпульсної характеристики

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) z^{-n}. \quad (5.18)$$

Враховуючи, що $H(z) = Y(z)/X(z)$ дорівнює відношенню z -перетворень двох сигналів (вихідного і вхідного), її називають *системною*

або передавальною функцією. Вона відіграє дуже важливу роль при побудові дискретних цифрових систем.

5.4. Цифрові фільтри. Класифікація цифрових фільтрів

Дискретний (цифровий) фільтр – це довільна система (програмний чи апаратний засіб) обробки дискретного сигналу, що має властивості лінійності та стаціонарності. Будь-який фільтр має певну частотну характеристику. Щоб вона була не тривіальною, тобто щоб коефіцієнти передачі фільтру на різних частотах були різними, вихідний сигнал фільтру $y(n)$ повинен залежати від кількох відліків вхідного сигнал $x(n)$. Отже дискретний фільтр повинен мати пам'ять.

Взаємозв'язок між вхідним $x(n)$ і вихідним $y(n)$ дискретними сигналами визначається різницеvim рівнянням [10]

$$y(n) = \sum_{k=0}^{N-1} b_k x(n-k) - \sum_{m=0}^{M-1} a_m y(n-m). \quad (5.19)$$

Тут границі сум N й M та величини $y(n)$ є коефіцієнтами (параметрами) фільтра, причому коефіцієнти a_m і b_k можуть бути як константами, так (у більш загальному випадку) відліками функцій, що залежать від дискретного часу n .

Сигнали $x(n)$ і $y(n)$ можуть бути як дійсними, так і комплексними. Рівняння виду (5.19) можна розглядати як алгоритм обчислення $y(n)$, тобто алгоритм роботи цифрового фільтру.

Для забезпечення лінійності та стаціонарності, операції що виконуються фільтром повинні обмежуватись лише додаванням та множенням на константу.

За призначенням цифрові фільтри умовно поділяються на два великі класи:

- 1) частотні фільтри;
- 2) функціональні фільтри.

Частотні фільтри призначені для цілеспрямованої зміни частотного спектра вхідного сигналу. Вихідними даними для проектування таких фільтрів є бажані частотні характеристики, що задаються. Реальні частотні характеристики цифрових фільтрів мають бути досить близькими до бажаного. Ступінь близькості бажаної і реальної частотних характеристик

визначається деякою кількісною мірою і є показником якості частотних фільтрів.

До частотних фільтрів належать:

- селективні (частотно-вибіркові) фільтри;
- коригуючі фільтри;
- перетворювачі Гільберта.

У частотно-вибіркових фільтрах повністю приглушуються частотні складові у певній смузі частот, яка називається смугою затримки, і зберігаються незмінними частотні складові в іншій смузі частот, що називається смугою пропускання. Між смугами пропускання і затримки можливий діапазон частот, характер поведінки складових у яких не залежить від вихідного сигналу (перехідна смуга).

За аналогією з аналоговими фільтрами, в залежності від положення зазначених смуг і їхньої кількості розрізняють наступні частотно-вибіркові фільтри:

- фільтри низьких частот;
- фільтри високих частот;
- смугові фільтри;
- режекторні фільтри.

Функціональні фільтри використовуються для виконання більш складних операцій над сигналами, що напряду можуть і не бути пов'язаними зі зміною частотного спектра вхідного сигналу. Вхідний сигнал у таких фільтрах, як правило, представляється у вигляді адитивної суміші корисного сигналу $u(n)$ і випадкового шуму $\xi(n)$ з відомою кореляційною функцією або функцією спектральної щільності потужності $S(\omega)$.

До функціональних фільтрів можна віднести:

- фільтри узгодження, що оцінюють ступінь відповідності корисної складової вхідного сигналу деякому еталонному сигналові на фоні шумів;
- екстраполюючі фільтри, що обчислюють майбутнє значення корисного сигналу в умовах дії випадкових завад.

Оскільки повне придушення випадкових завад неможливе, то оцінки, що отримуються на виході функціональних фільтрів, носять статистичний

характер. У якості їхньої кількісної міри зазвичай використовується дисперсія.

Спрощену блок-схему цифрового фільтру реального часу з аналоговими входом та виходом наведено на рис. 5.3.



Рис. 5.3. Спрощена блок-схема цифрового фільтру

Відповідно до схеми періодично робиться вибірка певних значень вузькосмугового аналогового сигналу, яка потім конвертується в набір цифрових вибірок $x(n)$, $n=0, 1, \dots$

Цифровий процесор виконує фільтрацію, відображаючи вхідну послідовність $x(n)$ у вихідну послідовність $y(n)$ відповідно до обчислювального алгоритму фільтру. ЦАП конвертує відфільтрований цифровим образом вихід в аналогове значення, яке потім проходить аналогову фільтрацію для згладжування та видалення небажаних високочастотних компонентів.

Важливу роль в цифровій обробці сигналів відіграє саме цифровий фільтр. Порівняно з аналоговими фільтрами, вони переважають в багатьох областях (наприклад: стиснення даних, біомедична обробка сигналів, обробка мови, зображень, передача даних, цифрове аудіо тощо), оскільки мають ряд наступних переваг:

- цифрові фільтри можуть мати характеристики, отримати які на аналогових фільтрах неможливо, наприклад дійсно лінійну фазову характеристику;
- на відміну від аналогових, продуктивність цифрових фільтрів не залежить від зміни середовища, наприклад від коливання температури;
- якщо фільтр розроблений на основі програмованого процесора, його частотна характеристика може настраюватись автоматично, тому такі процесори широко використовуються в адаптивних фільтрах;
- один цифровий фільтр може оброблювати декілька вхідних сигналів або каналів без дублювання апаратних блоків;

- як відфільтровані, так і невідфільтровані дані можна зберігати для наступного використання;

- на практиці точність, яку можна отримати при використанні аналогового фільтру, обмежена; наприклад, затухання в смузі затримки не можна підняти вище (60 - 70) дБ (якщо використовувати стандартні аналогові компоненти). Точність цифрових фільтрів обмежена лише довжиною слова.

- цифрові фільтри можуть застосовуватись при дуже низьких частотах, характерних, наприклад, для багатьох біомедичних додатків, де застосовувати аналогові фільтри не практично. Окрім того, дані фільтри можна використовувати у широкому діапазоні частот, для чого достатньо просто змінювати частоту дискретизації.

- Однак, в порівнянні з аналоговими, цифрові фільтри мають ряд недоліків:

- *обмежена швидкість*. Максимальна ширина смуги сигналів, які в реальному часі здатні опрацьовувати ці фільтри значно вужча, ніж у аналогових фільтрів. В додатках реального часу процес перетворення «аналоговий-цифровий-аналоговий» вводить обмеження за швидкістю на продуктивність цифрових фільтрів. Найвищу частоту дискретизації, з якою може працювати фільтр, обмежує час конвертації АЦП та час встановлення сигналу ЦАП. Окрім того, швидкість роботи цифрових фільтрів залежить від швидкості роботи цифрового процесора та числа арифметичних операцій, які необхідно виконати в алгоритмі фільтрації, і збільшується, коли характеристики фільтру стають більш стиснутими;

- *вплив кінцевої розрядності*. Цифрові фільтри піддаються зашумленню АЦП, що виникає при квантуванні неперервного сигналу та шуму округлення, який вводить при розрахунках. При застосуванні рекурсивних фільтрів високий порядок накопленого шуму округлення може призвести до нестійкості фільтру;

- *значний час розробки та впровадження*. Розробка та впровадження цифрових фільтрів, особливо впровадження апаратного забезпечення, можуть виконуватись набагато довше, ніж подібні процедури для аналогових фільтрів. Однак, одного разу розроблене апаратне або

програмне забезпечення може використовуватись в інших задачах цифрової обробки сигналів з незначною модифікацією або і зовсім без неї.

Цифрові фільтри також прийнято поділяти на два класи: рекурсивні (РФ) та не рекурсивні (НРФ). Якщо в рівнянні (5.19) хоча б один коефіцієнт a_m є відмінним від нуля, то фільтр називається *рекурсивним*.

Якщо ж у (5.19) всі коефіцієнти $a_m=0$, то фільтр, що реалізує такий алгоритм, називається *нерекурсивним* або *трансверсальним*. Для нього рівняння (5.19) спрощується до наступного виду:

$$y(n) = \sum_{k=0}^{N-1} b_k x(n - k). \quad (5.20)$$

Очевидно, що НРФ є системою без зворотного зв'язку, а РФ – системою зі зворотними зв'язками.

Розглянемо більш детально обидва класи цифрових фільтрів.

5.5. Нерекурсивні фільтри

В загальному випадку (формула (5.19)) при розрахунку чергового вихідного відліку $y(n)$ використовується інформація двох типів: певна кількість відліків вхідного сигналу та певна кількість попередніх відліків вихідного сигналу. Зрозуміло, що хоча б один відлік вхідного сигналу повинен брати участь при розрахунках, в іншому випадку вихідний сигнал не буде залежати від вхідного. На відміну від цього, попередні відліки вихідного сигналу можуть і не використовуватись, і рівняння фільтрації в цьому випадку прийме вигляд наведений в (5.20).

Кількість попередніх відліків ($N-1$), що використовуються при розрахунках, буде визначати порядок фільтру.

Структурна схема, реалізації алгоритму (5.20) наведена на рис. 5.4 [10]. Деяка кількість попередніх відліків вхідного сигналу зберігаються в комірці пам'яті, які утворюють дискретну лінію затримки. Ці відліки множаться на коефіцієнти b_k і додаються, формуючи вихідний відлік $y(n)$.

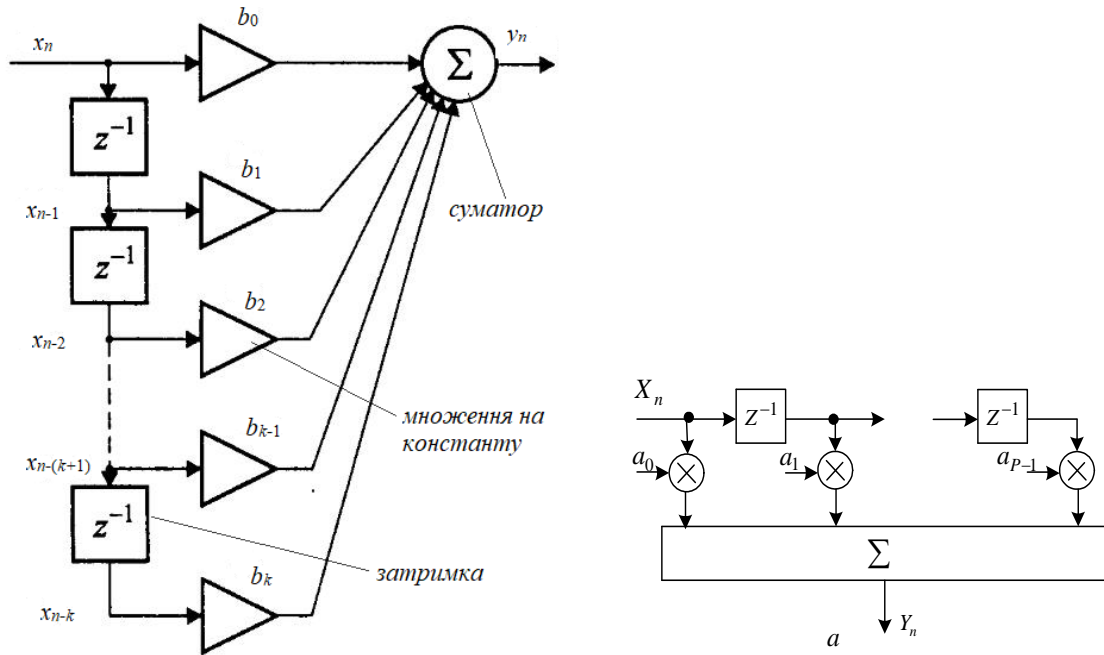


Рис. 5.4. Структура нерекурсивного фільтру

Відповідно до властивостей z -перетворення, затримка дискретної послідовності на один такт (так звана *одинична затримка*) відповідає множенню її z -перетворення на z^{-1} . Тому елементи пам'яті, що здійснюють таку затримку, позначаються на структурній схемі як « z^{-1} » і утворюють регістр зсуву.

Оскільки при розрахунках не використовується попередній відлік вихідного сигналу, в схемі відсутній зворотний зв'язок. Тому такі фільтри називаються нерекурсивні (*nonrecursive*). Також застосовують термін «трансверсальний фільтра» (від англ. *transversal* - поперечний).

Системну функцію НРФ що відповідає рівнянню (5.20) можна представити у наступному вигляді

$$H_{\text{НРФ}}(z) = \sum_{k=0}^{N-1} b_k z^{-k}. \quad (5.21)$$

Визначимо імпульсну характеристику НРФ. Для цього в рівняння (6.20) треба підставити одиничний імпульс $x_0(n)$ в якості вхідного сигналу:

$$h(n) = \sum_{k=0}^{N-1} b_k x_0(n-k). \quad (5.22)$$

Так як відлік $x_0(n-k)$ дорівнює нулю для всіх n , окрім $n=k$, коли цей відлік дорівнює одиниці, то отримуємо досить простий результат:

$$h(k) = b_k, \quad (5.23)$$

тобто коефіцієнти b_k є відліками імпульсної характеристики фільтру.

Це легко пояснити за допомогою рис. 5.4. При подачі на вхід системи одиничного імпульсу він буде переміщатись по лінії затримки, помножуючись на коефіцієнти b_0, b_1, b_2, \dots та проходить на вихід пристрою (адже всі інші вхідні сигнали суматора при цьому дорівнюють нулю).

Імпульсну характеристику цифрового фільтру можна сприймати як результат дискретизації неперервної імпульсної характеристики аналогового фільтру-прототипу. Можливі два варіанти кількості відліків $h(n)$ – кінцева та нескінченна.

Очевидно, що в реальному цифровому фільтрі лінія затримки містить скінченне число елементів, тому імпульсна характеристика НРФ також є скінченною за тривалістю. Тому НРФ відносяться до фільтрів з *кінцевою імпульсною характеристикою* (КІХ-фільтри, *finite impulse response*, FIR).

Із (5.23) витікає, що коефіцієнти КІХ фільтра та його імпульсна характеристика є синонімами.

З врахуванням формули (5.22) запишемо

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k). \quad (5.24)$$

Отримане рівняння є рівнянням згортки для цифрових КІХ фільтрів. Процес, що описується (5.24), полягає в певній кількості перемножень і наступному підсумовуванні добутоків.

Стандартною формою запису згортки (5.24) є

$$y(n) = h(k) * x(n),$$

де символ * означає згортку.

Для КІХ фільтрів ДПФ згортки імпульсної характеристики фільтру та вхідного сигналу дорівнює добутку спектру вхідного сигналу та ДПФ імпульсної характеристики. Цей процес для $N=5$ проілюстровано на рис. 5.5 [5].

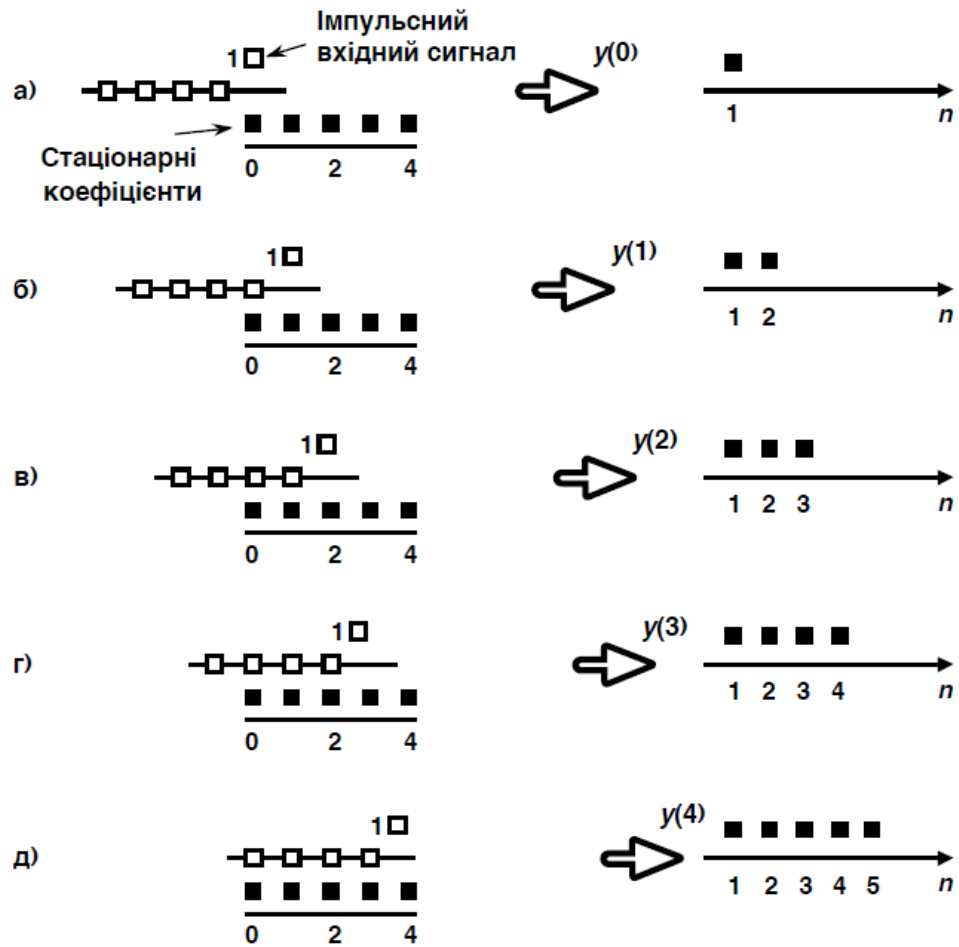


Рис. 5.5. Згортка коефіцієнтів фільтра та вхідного імпульсу для отримання імпульсної характеристики фільтра: а - одиничний відлік імпульсу вирівнено з першим коефіцієнтом фільтра, індекс $n = 0$; б - імпульс зсунуто вправо, індекс $n = 1$; в - індекс $n = 2$; г - індекс $n = 3$; д - індекс $n = 4$

Тут коефіцієнти фільтра $h(k)$ позначено чорними квадратами, відліки вхідного сигналу – порожніми квадратами (значення відліків є однаковими між собою, перший відлік зміщено вгору для наочності).

На кожному наступному рисунку (а) – (е) вхідний сигнал зміщується направо на один крок. В правій частині чорними квадратами показані добутки $h(k)x(n - k)$. Якщо скласти для кожного пункту ці добутки, то отримаємо результати обчислення по формулі (А): (а) – $y(1)$, (б) - $y(2)$, (в) - $y(3)$, (д) - $y(4)$, (е) - $y(5)$.

У зв'язку з відсутністю зворотного зв'язку будь-який НРФ є стійким, оскільки якими б не були початкові умови (тобто відліки, що зберігаються в лінії затримки), при відсутності сигналу на вході ($x(n)=0$) вихідний сигнал

(вільні коливання) не буде дорівнювати нулю протягом $(N-1)$ тактів, необхідних для очищення лінії затримки.

Простота аналізу та реалізації, а також наглядний зв'язок коефіцієнтів фільтру з відліками його імпульсної характеристики та абсолютна стійкість призвели до того, що НРФ знайшли широке застосування на практиці. Однак, для отримання кращих частотних характеристик (наприклад, смугових фільтрів з крутою прямокутною АЧХ) необхідні НРФ досить високого порядку – до декількох сотень і навіть тисяч.

5.6. Рекурсивні фільтри

Якщо рівняння фільтрації містить як вхідні, так і вихідні відліки, то його робота описується формулою (5.19). Для реалізації такого фільтру в схему, що наведена на рис. 5.4, необхідно додати другу лінію затримки – для зберігання вихідних відліків $y(n-k)$. Отримана при цьому структура фільтру наведена на рис. 5.6 [10].

Оскільки при розрахунках використовуються попередні відліки вихідного сигналу, в схемі присутній зворотний зв'язок. Кількість попередніх вхідних та вихідних відліків може не співпадати. В такому випадку порядок фільтру визначається максимальним із значень $(N-1)$ та $(M-1)$.

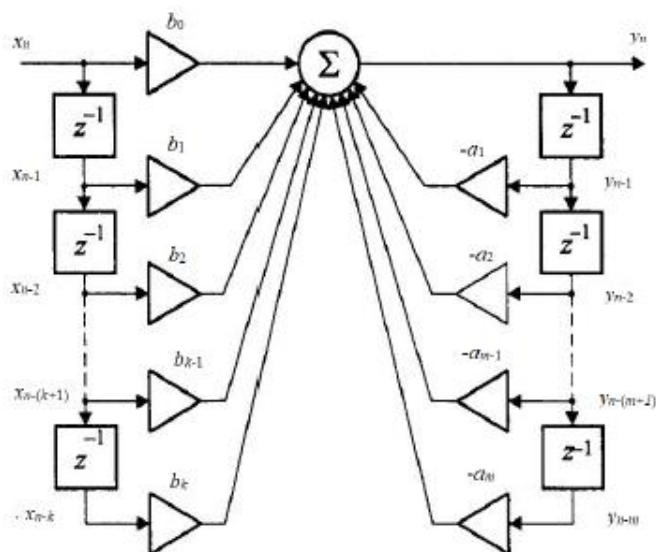


Рис. 5.6. Структура рекурсивного фільтру, пряма реалізація

Наявність зворотніх зв'язків дозволяє отримати в таких фільтрах нескінченну імпульсну характеристику (НІХ). Тому РФ відносяться до класу НІХ фільтрів (*infinite impulse response*, ІІР). Відповідно до цієї причини, РФ можуть бути нестійкими і перетворюються на генератори.

Імпульсну характеристику рекурсивного фільтра розраховувати складніше, ніж для не рекурсивного. При надходженні на вхід одиничного імпульсу він множиться на b_0 і проходить на вихід. Таким чином:

$$h(0) = b_0. \quad (5.25)$$

Далі вхідний одиничний імпульс потрапляє у вхідну лінію затримки, а вихідний відлік, рівний b_0 – у вихідну лінію затримки. В результаті другий відлік імпульсної характеристики буде формуватись наступним чином:

$$h(1) = b_1 + a_1 h(0) = b_1 + b_0 a_1. \quad (5.26)$$

Наступний відлік буде визначатись як

$$h(2) = b_2 + a_2 h(0) + a_1 h(1) = b_2 + b_0 a_2 + a_1 (b_1 + b_0 a_1) = b_2 + b_1 a_1 + b_0 a_2 + b_0 a_1^2.$$

Таким чином, по мірі того, як вихідна лінія затримки заповнюється відліками імпульсної характеристики, складність аналітичних формул швидко наростає. Не дивлячись на складність, НІХ фільтри використовуються досить часто завдяки своїй ефективності.

Вони вимагають набагато менше перемножень на один відлік для формування потрібної частотної характеристики. Це означає, що такі фільтри можуть бути дуже швидкими і можуть працювати в реальному часі, на більших, ніж у КІХ фільтрів частотах дискретизації.

Нажаль немає простих методів обчислення коефіцієнтів НІХ фільтрів a_k, b_k по імпульсній характеристиці. Стандартні методи проектування НІХ фільтрів розділяються на три групи: методи інваріантного перетворення імпульсної характеристики, методи білінійного перетворення, оптимізаційні методи. Ці методи засновані на використанні z -перетворення.

Існує досить велике число різних форм реалізації рекурсивних і нерекурсивних ЦФ. Розглянемо найбільш поширені з них.

Структурна схема, що наведена на рис. 5.6. називається *прямою формою* реалізації РФ (*direct form I*) і не є єдиною можливою. Пряма форма відповідає безпосередній реалізації різницевого рівняння (5.19).

Розділивши суматор на схемі рис. 5.6 на два окремих – для рекурсивної

та не рекурсивної частини фільтру (рис. 5.7, а), отримаємо два послідовно з'єднаних фільтри, один з яких є не рекурсивним, а інший – містить лише рекурсивну частину.

Так як результат послідовного проходження сигналу через ряд лінійних стаціонарних пристроїв не залежить від послідовності їх з'єднання, можемо поміняти місцями дві половинки нашого фільтру (рис. 5.7, б).

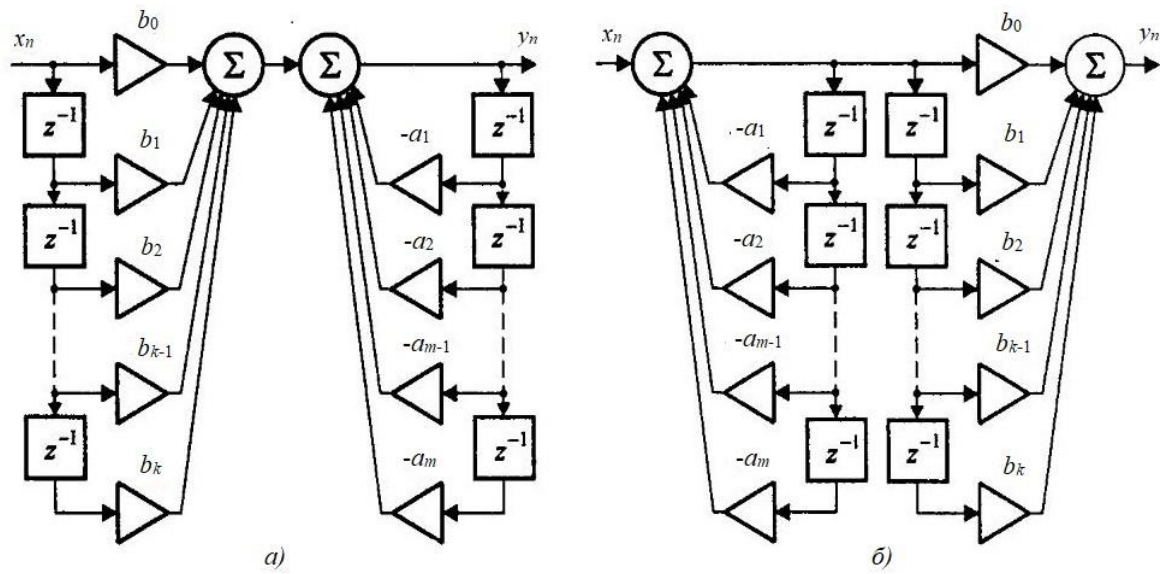


Рис. 5.7. Перестановка рекурсивної та нерекурсивної частини фільтру

Видно, що в обидві лінії затримки подається один і той же сигнал, а отже вони будуть мати однакові набори відліків. Це дозволить об'єднати лінії затримки. Отримана в результаті схема, зображена на рис. 5.8, називається канонічною формою реалізації рекурсивного фільтру (*canonic form* або *direct form II*).

При канонічній реалізації застосовується загальна лінія затримки, що зменшує число необхідних комірок пам'яті. Однак при цьому абсолютні значення відліків, що проходять лінію затримки, можуть суттєво перевищувати амплітуду вхідного та вихідного сигналів. Це призводить до необхідності збільшувати розрядність представлення чисел в лінії затримки в порівнянні з розрядністю вхідного і вихідного сигналів, що ускладнює реалізацію пристрою.

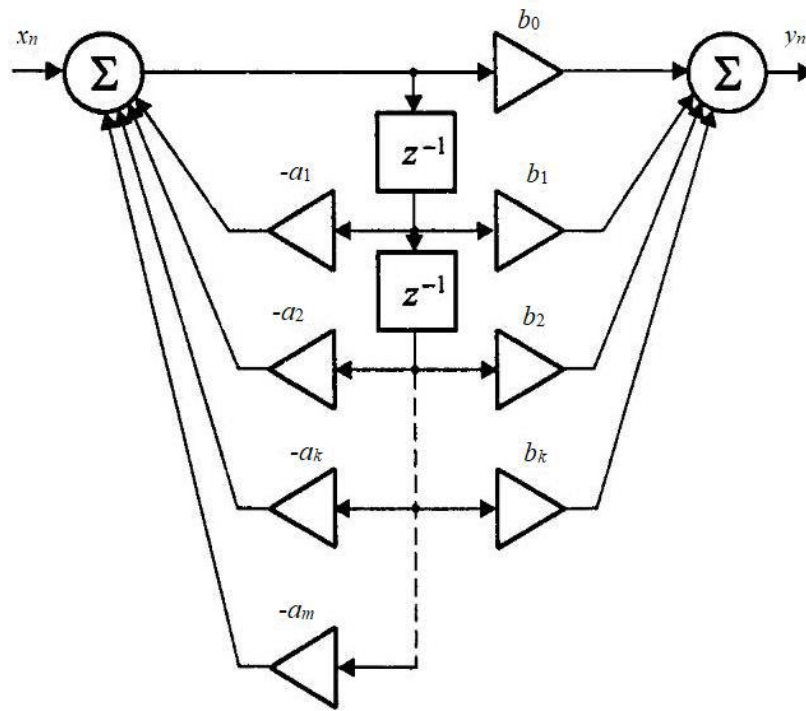


Рис. 5.8. Рекурсивний фільтр, канонічна реалізація

При прямій реалізації в лініях затримки зберігаються безпосередні відліки вхідного та вихідного сигналів, тобто підвищення розрядності лінії затримки не потрібне. Єдиним елементом, що потребує підвищення розрядності, в даному випадку є суматор, і це враховується в архітектурі мікропроцесорів, що спеціально призначенні для обробки сигналів в реальному часі.

Існують також інші структури НІХ фільтрів. Причому всі форми реалізації фільтрів при тих самих вхідних даних і нескінченній розрядності представлення чисел дають абсолютно однакові результати, тому що отримані шляхом еквівалентних математичних перетворень одного й того ж вихідного різницевого рівняння (6.19).

Однак, при обмеженій розрядності сітки представлення чисел, що завжди має місце в реальних фільтрах, ці форми приведуть до різних результатів, тому що відрізняються механізмом перетворення похибок округлення.

5.7. Порівняння типів ЦФ

В залежності від імпульсної характеристики цифрові фільтри прийнято поділяти на два класи КІХ-фільтри та НІХ-фільтри. Як вже зазначалось, всі НРФ є КІХ-фільтрами, а майже всі РФ (за винятком деяких) є НІХ-фільтрами.

Вибір між КІХ та НІХ-фільтрами залежить від відносних переваг обох фільтрів:

1. КІХ-фільтри можуть мати строго лінійну фазову характеристику. Відповідно, фільтр не вносить фазового спотворення в сигнал, що важливо в багатьох сферах, наприклад при передачі даних, біомедицині, цифровій аудіо обробці або обробці зображень. Фазова характеристика НІХ-фільтра нелінійна, особливо на краях смуг.

2. КІХ-фільтри завжди стійкі, гарантувати стійкість НІХ-фільтрів вдається не завжди.

3. Для реалізації фільтрів застосовується обмежена кількість бітів. Практичні наслідки цього (шум округлення та помилки квантування) менш істотні для КІХ-фільтрів, ніж для НІХ-фільтрів.

4. Щоб отримати кінцеву послідовність імпульсної характеристики за допомогою фільтрів з різкими зрізами характеристик, необхідно більше коефіцієнтів, ніж для отримання НІХ. Відповідно для реалізації необхідної амплітудної характеристики з КІХ необхідно більше розрахунків, ніж при реалізації НІХ. Однак, ефективність КІХ-реалізації можна значно підвищити, застосувавши швидке перетворення Фур'є та обробку при декількох швидкостях.

5. Аналогові фільтри легко перетворити в еквівалентні цифрові НІХ-фільтри, що задовольняють подібним вимогам. Для отримання КІХ-фільтрів таке перетворення неможливе, оскільки для них не існує аналогових прототипів. Однак, отримувати довільні частотні характеристики на КІХ-фільтрах легше.

6. НІХ-фільтри рекурентні. Це означає, що, пропустивши через фільтр один і той самий сигнал, але зі «зворотним ходом часу», можна отримати, різні результати. Якщо для мови часова анізотропія природня, то,

наприклад, для зображень, вже ні. Тому НІХ-фільтри мають ряд обмежень на застосування.

Враховуючи вище зазначене, загальний принцип вибору між КІХ та НІХ можна сформулювати наступним чином:

- використовувати НІХ, якщо єдиною важливою вимогою є характеристика з різким зрізом і висока пропускну спроможність, оскільки НІХ-фільтри (особливо ті, в яких застосовується еліптичні характеристики) потребують меншого числа коефіцієнтів, а отже і розрахунків;

- застосовувати КІХ, якщо число коефіцієнтів фільтрів не дуже велике, наприклад, якщо необхідно, щоб фазові спотворення були відсутні або були малими. Окрім того, можна додати, що архітектура процесорів цифрової обробки сигналів більш пристосована до КІХ-фільтрації, й більш того, деякі з них спеціально розроблені для фільтрів даного типу.

Питання для самоперевірки

1. Визначення лінійних і нелінійних систем.
2. Що таке різницева система?
3. Фізичний зміст інтеграла Дюамеля.
4. Що таке імпульсна характеристика системи?
5. Надайте характеристику типових аналогових фільтрів. Чим відрізняються ідеальні та реальні фільтри?
6. Що називають цифровим фільтром?
7. Класифікація цифрових фільтрів.
8. В чому полягає сенс нерекурсивного фільтру?
9. Які відмінності в структурах нерекурсивних та рекурсивних фільтрів?
10. В чому полягають відмінності різних типів цифрових фільтрів?

РОЗДІЛ 6. МОДУЛЯЦІЯ ТА ДЕМОДУЛЯЦІЯ СИГНАЛІВ

При створенні каналів зв'язку інформаційно-вимірювальних систем для ефективного передавання низькочастотних сигналів на великі відстані виникає потреба у перенесенні їх на більш високі. При цьому також дуже часто в одному й тому ж каналі зв'язку необхідно передавати декілька сигналів одночасно.

Якщо різні сигнали займають смуги частот, що не перекриваються, можливим є, так зване, частотне розділення. Частіше за все вимагається, щоб випромінюваний сигнал був *вузькосмуговим*, тобто ефективна ширина його спектру була набагато менше його центральної частоти $\Delta f \ll f_0$.

Тому з метою передавання вимірювальний сигнал перетворюється так, щоб вимоги до смуги частот каналу передавання виконувались, а сам вхідний сигнал можна було б відновити. Подібні перетворення застосовуються також в аналізаторах спектру сигналів.

6.1. Модуляція сигналів

Вирішення вказаної задачі досягається шляхом використання модуляції, тобто процесу зміни в часі одного або декількох параметрів сигналу-переносника з метою передавання інформації каналом зв'язку. Суть модуляції полягає в наступному. Формується деяке електромагнітне коливання – *коливання-носії* (carrier) (зустрічаються також терміни *несівне коливання*, *носійне коливання*, *сигнал-переносник*).

Певний параметр цього коливання змінюється в модуляторі пропорційно сигналу, який треба передати. Змінений сигнал називають модульовальним, а результуюче коливання з параметрами, що змінюються в часі, – модульованим коливанням. Зворотний процес – виділення модульовального сигналу з модульованого коливання – називається демодуляцією.

На рисунку 6.1 представлено функціональну схему модулятора, де $u(t)$ – вхідний модульовальний сигнал (сигнал, що модулює); $s(t)$ – вихідний сигнал (модульований сигнал); $s_0(t)$ – коливання-носії. В процесі *модуляції*

параметри коливання-носія $s_0(t)$ змінюються під впливом вхідного сигналу $u(t)$ для отримання модульованого сигналу $s(t)$.

Коливання-носії частіш за все є гармонічним

$$s_0(t) = A_0 \cos(\omega_0 t + \varphi_0) = A_0 \cos \theta(t), \quad (6.1)$$

де A_0 – амплітуда коливання, $\omega_0 = 2\pi f_0$ – кутова частота, φ_0 – початкова фаза, $\theta(t)$ – повна фаза.

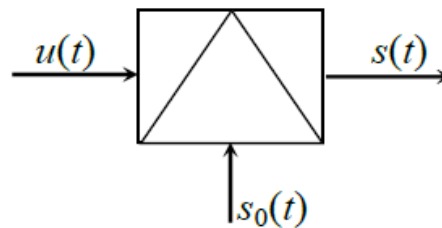


Рис. 6.1. Функціональна схема модулятора

Основними видами модуляції є амплітудна (АМ, amplitude modulation) і кутова (КМ, angle modulation). Якщо змінним параметром є амплітуда коливання-носія (s_0), то сигнал буде *амплітудно-модульований* ($s_{AM}(t)$). При зміні повної фази $\theta(t)$ отримаємо сигнал з *кутовою модуляцією* ($s_{KM}(t)$). Кутова модуляція поділяється на два тісно пов'язані між види: частотна модуляція (ЧМ, frequency modulation) і фазова модуляція (ФМ, phase modulation). Сигнали з такими видами модуляції позначатимемо відповідно як $s_{ЧМ}(t)$ і $s_{ФМ}(t)$. На практиці зустрічаються й змішані види модуляції: амплітудно-фазова, амплітудно-частотна тощо [11].

При *амплітудній модуляції* у відповідності до модульовального сигналу змінюється амплітуда коливання-носія

$$s_{AM}(t) = A(t) \cos(\omega_0 t + \varphi_0).$$

Для реалізації АМ до модульовального сигналу зазвичай додають постійну складову (це робить його однополярним і дозволяє запобігти спотворенню обвідної двополярного сигналу при його демодуляції)

$$A(t) = A_0(1 + m s_M(t)),$$

де m – коефіцієнт амплітудної модуляції (коефіцієнт глибини модуляції), практично його значення приймають в межах (0,4 - 0,6); $s_M(t)$ – сигнал, що модулює.

Остаточно АМ сигнал можна записати в наступному виді

$$s_{AM}(t) = A_0(1 + ms_M(t))\cos(\omega_0 t + \varphi_0). \quad (6.2)$$

Спектральну структуру АМ сигналу зручніше дослідити для випадку *однотональної* АМ, коли модулювальний сигнал є гармонічним

$$s_M(t) = A_M \cos(\Omega t + \Phi_0),$$

з частотою $\Omega \ll \omega_0$.

В цьому випадку

$$s_{AM}(t) = A_0(1 + m\cos(\Omega t + \Phi_0))\cos(\omega_0 t + \varphi_0). \quad (6.3)$$

Коефіцієнт модуляції $m = A_M/A_0$ може бути обчислений за результатами вимірювання (наприклад, за допомогою осцилографа) максимальної A_{\max} і мінімальної A_{\min} амплітуд сигналу: $m = (A_{\max} - A_{\min}) / (A_{\max} + A_{\min})$.

В окремих випадках вимірюють відносні коефіцієнти модуляції вгору $m_{\text{в}} = (A_{\max} - A_0) / A_0$ або вниз $m_{\text{н}} = (A_0 - A_{\min}) / A_0$.

На рис. 6.2 приведені епюри напруг для різних значень коефіцієнта амплітудної модуляції, які дають уяву про вигляд сигналу при порівняно неглибокій (б), глибокій модуляції (в) та при його перемодуляції (г).

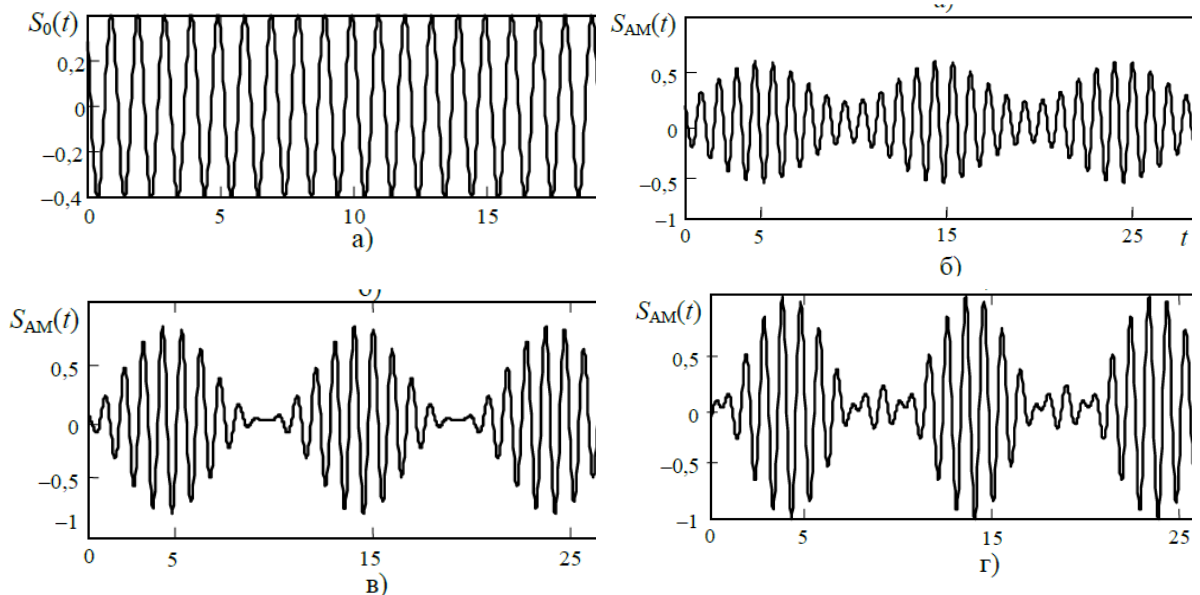


Рис. 6.2. Однотональний АМ сигнал:

а - $m=0$ (коливання-носії); б - $m=0,5$; в - $m=1$; г - $m=1,5$

Якщо коефіцієнт амплітудної модуляції лежить в діапазоні (0 – 1), обвідна модульованого сигналу повторює форму модулюючого сигналу

$s_M(t)$ (рис. 6.2, б, в).

Перемодуляція має місце при $m > 1$, при цьому амплітудна обвідна спотворюється (рис. 6.2, г). Проте і такий режим в деяких випадках використовується на практиці.

З метою аналізу спектрального складу такого коливання розкриємо дужки у виразі (6.3) й виконаємо тригонометричні перетворення

$$s_{AM}(t) = A_0 \cos(\omega_0 t + \varphi_0) + A_0 m \cos(\Omega t + \Phi_0) \cos(\omega_0 t + \varphi_0) = A_0 \cos(\omega_0 t + \varphi_0) + \frac{A_0 m}{2} \cos((\omega_0 + \Omega)t + \varphi_0 + \Phi_0) + \frac{A_0 m}{2} \cos((\omega_0 - \Omega)t + \varphi_0 - \Phi_0). \quad (6.4)$$

Результат перетворення показує, що однотональний АМ сигнал складається з трьох гармонічних складових. Одна з них є коливанням-носієм з частотою ω_0 , а дві інші – бічні коливання з частотами $(\omega_0 - \Omega)$ і $(\omega_0 + \Omega)$ (їх називають нижньою і верхньою бічними частотами). Амплітуда коливання-носія дорівнює A_0 і не залежить від рівня модулюючого сигналу. Амплітуди ж коливань бічних частот дорівнюють $A_0 m/2$ і є пропорційними коефіцієнту модуляції.

Для коливання з частотою $(\omega_0 + \Omega)$ початкові фази коливання-носія і модулюючого сигналу додаються, а для коливання з частотою $(\omega_0 - \Omega)$ – віднімаються. Амплітудний і фазовий спектри однотонального АМ сигналу показані на рис. 6.3. Цей спектр є дискретним, але в загальному випадку він не є розкладанням в ряд Фур'є. Це має місце лише тоді, коли частоти спектральних складових мають загальний дільник, тобто співвідношення ω_0/Ω є раціональним дробом.

Видно також, що ширина спектру однотонального АМ сигналу в два рази перевищує частоту модулюючого сигналу: $\Delta\omega = 2\Omega$.

Амплітуда модульованого сигналу змінюється в межах від мінімального значення

$$s_{AMmin} = A_0 - \frac{mA_0}{2} - \frac{mA_0}{2} = A_0(1 - m)$$

до максимального значення

$$s_{AMmax} = A_0 + \frac{mA_0}{2} + \frac{mA_0}{2} = A_0(1 + m).$$

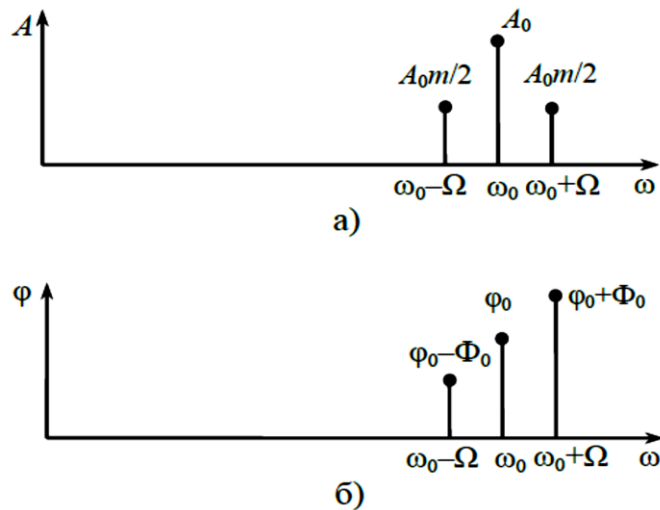


Рис. 6.3. Спектр однотонового АМ сигналу: а) амплітудний; б) фазовий

Зі зміною амплітуди S_{AM} змінюється й середня потужність сигналу:

- поблизу до максимальної амплітуди потужність

$$P_{\max} = \frac{S_{AM\max}^2}{2} = \frac{A_0^2(1+m)^2}{2} = P_0(1+m)^2;$$

- в околиці мінімальної S_{AM} потужність

$$P_{\min} = \frac{S_{AM\min}^2}{2} = \frac{A_0^2(1-m)^2}{2} = P_0(1-m)^2.$$

Максимальна потужність при $m = 1$, становить

$$P_{\max} = 4P_0.$$

Тепер знайдемо середню потужність АМ сигналу. Оскільки гармонічні коливання різних частот є некорельованими, то їх потужності можна складати. Застосувавши сказане до подання однотонового АМ сигналу (6.4) отримаємо співвідношення

$$P_{\text{сеп}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} S_{AM}^2(t) dt = \frac{A_0^2}{2} + \frac{A_0^2 m^2}{4}.$$

Перший доданок не залежить від коефіцієнта модуляції і є потужністю немодульованого коливання-носія. Корисна потужність, яка зосереджена в бічних коливаннях, представлена другим доданком.

Амплітудні модулятори частіш за все будуються за принципом перетворення спектру суми двох сигналів (модулюючого та гармонічного коливання-носія) в безінерційному нелінійному пристрої. Цей принцип можна реалізувати, наприклад, із застосуванням нелінійного

транзисторного підсилювача.

Історично АМ була першим практично освоєним видом модуляції. Однак низький ККД і ширина спектра, що вдвічі перевищує ширину спектра модулюючого сигналу, обмежили сфери застосування АМ у зв'язку. В даний час АМ застосовується для радіомовлення на порівняно низьких частотах (в діапазонах довгих, середніх і коротких хвиль) і для передавання зображення в телевізійному мовленні.

З метою поліпшення характеристик АМ розроблено кілька її модифікацій. Наприклад, очевидно, що для підвищення ККД амплітудної модуляції доцільно видалити «зайве» коливання-носії, відмовившись від додавання постійної складової до модулюючого сигналу. Такий спосіб називається АМ з придушеною частотою-носієм (АМ-ПН, *amplitude modulation with suppressed carrier, AM-SC*) або балансною АМ:

$$s_{AM}(t) = s_M \cos(\omega_0 t + \varphi_0).$$

Ширина спектра АМ сигналу з придушеною частотою-носієм така ж, як у випадку звичайної АМ (адже придушена середня частота, а бічні частоти залишилися на місці). Таким чином, амплітудна модуляція з придушеною частотою-носієм має певні енергетичні переваги в порівнянні зі звичайною АМ. Однак цей спосіб модуляції не отримав широкого розповсюдження, що пов'язано з проблемами, які виникають при демодуляції сигналу. Крім того, ширина спектра сигналу при двосмуговій АМ з придушеною частотою-носієм залишається рівною подвоєній частоті модулюючого сигналу.

Однак, оскільки спектри двох бічних смуг АМ сигналу є дзеркальним відображенням один одного, вони несуть одну й ту саму інформацію. Тому одну з бічних смуг можна видалити. Отримана *модуляція називається односмуговою (single side band, SSB)*. Залежно від того, яка бічна смуга зберігається, говорять про односмугову модуляцію з використанням верхньої або нижньої бічної смуги. По суті, при односмуговій модуляції відбувається просто зсув спектра сигналу в околиці частоти-носія.

Очевидно, що ширина спектра односмугового сигналу дорівнює ширині спектра модулюючого сигналу. Таким чином, спектр односмугового сигналу виявляється в два рази вузчим, ніж при звичайній АМ.

При фазовій модуляції модулюючий сигнал визначає початкову фазу коливання-носія: $\varphi(t) = ks_M(t)$. Тому сигнал з фазовою модуляцією (ФМ, phase modulation, PM) має вигляд

$$s_{\text{ФМ}}(t) = A \cos(\omega_0 t + ks_M(t)),$$

де k – коефіцієнт пропорційності, $s_M(t)$ – закон модуляції.

Позначимо повну фазу коливання: $\Psi(t) = \omega_0 t + ks_M(t)$ і прийнемо, що $-1 \leq s_M(t) \leq 1$. При $s_M(t)=1$ коефіцієнт k дорівнює максимальному відхиленню (девіації) фази вгору і при $s_M(t)=-1$ – максимальному відхиленню фази вниз.

Кругова частота коливання за визначенням є швидкістю зміни фази. Для коливань з кутовою модуляцією вона змінюється у часі і називається *миттєвою частотою* (instantaneous frequency), яка є похідною від повної фази за часом

$$\omega(t) = \frac{d\Psi}{dt} = \omega_0 + k \frac{ds_M(t)}{dt}. \quad (6.5)$$

Отже, у разі фазової модуляції змінюється не тільки початкова фаза, але й миттєва частота коливання.

Відповідно, повна фаза може бути знайдена шляхом інтегрування миттєвої частоти

$$\Psi(t) = \int_{-\infty}^t \omega(\tau) d\tau. \quad (6.6)$$

У випадку *частотної модуляції* (ЧМ, frequency modulation, FM) модулюючий сигнал є лінійно пов'язаним з миттєвою частотою коливання

$$\omega(t) = \frac{d\Psi}{dt} = \omega_0 + ks_M(t).$$

Доданок у вигляді константи ω_0 робить коливання високочастотним. Повна фаза знаходиться інтегруванням $\omega(t)$

$$\Psi(t) = \omega_0 t + k \int_{-\infty}^t s_M(\tau) d\tau + \varphi_0.$$

де φ_0 – початкова фаза коливання в момент $t=0$, яка в загальному випадку може бути довільною.

Сам ЧМ сигнал має наступний вигляд

$$s_{\text{ЧМ}}(t) = A \cos \left[\omega_0 t + k \int_{-\infty}^t s_M(\tau) d\tau + \varphi_0 \right]$$

Як видно, початкова фаза коливання при частотній модуляції

змінюється пропорційно інтегралу від модулюючого сигналу:

$$\varphi(t) = k \int_{-\infty}^t s_M(\tau) d\tau + \varphi_0.$$

Фазову і частотну модуляції об'єднують під загальною назвою «кутова модуляція» саме, завдяки тому, що певній зміні початкової фази коливання відповідатиме також зміна миттєвої частоти (і навпаки). Це дозволяє зробити висновки:

- за формою коливання з КМ не можна визначити це сигнал з ФМ чи ЧМ, для цього потрібно також знати й форму модулюючого сигналу;

- якщо модулюючий сигнал пропустити через ідеальний диференціальний пристрій, з виходу якого подати на частотний модулятор, вийде ФМ сигнал (верхня гілка на рис. 6.4) і, навпаки, при проходженні модулюючого сигналу через ідеальний інтегровальний пристрій та далі через фазовий модулятор, отримаємо ЧМ сигнал (нижня гілка на рис. 6.4).

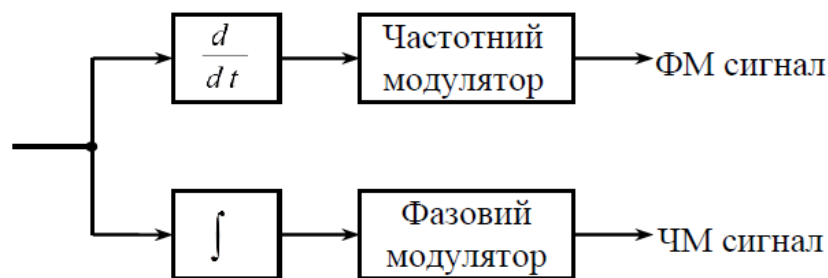


Рис. 6.4. Взаємозв'язок фазової і частотної модуляції

На відміну від АМ, у разі кутової модуляції отримати аналітичний вираз для спектра при довільному модулюючому сигналі не вдається. Навіть для модулюючого сигналу з суми гармонік з частотами Ω_1, Ω_2 в спектрі модульованого сигналу спостерігаються складові не тільки з частотами $(\omega_0 \pm \Omega_1), (\omega_0 \pm \Omega_2), (\omega_0 \pm 2\Omega_1), (\omega_0 \pm 2\Omega_2)$, але й складові з комбінаційними частотами $(\omega_0 \pm \Omega_1 \pm \Omega_2)$ з чотирма можливими сполученнями знаків перед Ω_1, Ω_2 .

Як вже згадувалось раніше, існують види модуляції, коли одночасно змінюються кілька параметрів коливання-носія, наприклад, одночасно можуть змінюватись амплітуда і початкова фаза гармонічного коливання

$$s(t) = A(t) \cos(\omega_0 t + \varphi_0(t)).$$

Таку модуляцію можна назвати амплітудно-фазовою і застосувати для одночасного передавання двох сигналів. Однак модулюючі сигнали в цьому випадку виявляються «нерівноправними», оскільки вони модулюють істотно різні параметри коливання-носія. Можна зробити ситуацію більш «симетричною», якщо розкрити косинус суми

$$s(t) = A(t) \cos[\varphi(t)] \cos \omega_0 t - A(t) \sin[\varphi(t)] \sin \omega_0 t.$$

В такому вигляді сигнал є сумою двох АМ коливань. Їх коливання-носії $\cos \omega_0 t$ і $\sin \omega_0 t$ мають зсув фази 90° по відношенню одна до одної, а от амплітуди дорівнюють відповідно

$$A(t) \cos[\varphi(t)] \quad \text{і} \quad A(t) \sin[\varphi(t)].$$

Позначивши амплітудні функції як $a(t)$ і $b(t)$, замість змінних амплітуди і початкової фази отримаємо пару модулюючих сигналів

$$s(t) = a(t) \cos \omega_0 t + b(t) \sin \omega_0 t.$$

Таке представлення називається *квадратурним* (quadrature), а даний спосіб модуляції – *квадратурною* модуляцією (КАМ, quadrature amplitude modulation, QAM). Модулюючі сигнали при цьому виявляються повністю рівноправними.

Для передачі цифрової інформації різноманітними каналами зв'язку у цифровому вигляді використовуються способи модуляції, які забезпечують передавання послідовності символів, кожен з яких є цілим числом з деякої фіксованої кінцевої множини.

Ці символи надходять від джерела інформації з частотою $f_T = 1/T$, де T – період їх слідування. Ця частота називається *символьною швидкістю* (symbol rate). На практиці часто використовується двійкова (binary) послідовність символів, коли кожне з чисел може мати одне з двох значень – 0 або 1.

Послідовність переданих символів є дискретним сигналом. Оскільки символи приймають значення з кінцевої множини, цей сигнал фактично є також квантованим. Тому, відповідно до визначень, введених на початку курсу, його можна назвати цифровим сигналом. Розглянемо перетворення цього цифрового сигналу в аналоговий модульований сигнал.

Типовий підхід до передавання дискретної послідовності символів полягає в наступному. Кожному з можливих значень символу зіставляється деякий набір параметрів коливання-носія. Ці параметри підтримуються постійними протягом інтервалу T , тобто до приходу наступного символу. Фактично це означає перетворення послідовності чисел $\{n_k\}$ в ступінчастий сигнал $s_n(t)$ з використанням кусково-неперервної інтерполяції

$$s_n(t) = f(n_k), kT \leq t < (k + 1)T,$$

де f – деяка функція перетворення.

Отриманий сигнал $s_n(t)$ далі використовується звичайним способом як модулюючий.

Спосіб модуляції, коли параметри коливання-носія змінюються стрибкоподібно, називається *маніпуляцією* (keying). Залежно від того, які саме параметри змінюються, розрізняють амплітудну (АМн), фазову (ФМн), частотну (ЧМн) і квадратурну (КАМн) маніпуляції. Крім того, при передаванні цифрової інформації може використовуватись коливання-носій відмінне за формою від гармонічного.

Так, при використанні в якості коливання-носія послідовності прямокутних імпульсів можливі амплітудно-імпульсна (АІМ), широтно-імпульсна (ШІМ) і фазово-імпульсна (ФІМ, РРМ – Pulse Position Modulation) види модуляції, де ШІМ і ФІМ є різновидами часово-імпульсної (ЧІМ, РТМ – Pulse Time Modulation).

Найпростіші приклади формування двохпозиційної (бінарної) маніпуляції показані на рис. 6.5 [11].

При *частотній маніпуляції* (frequency shift keying, FSK) кожному можливому значенню переданого символу протягом кожного символного інтервалу зіставляється своя частота (рис. 6.5, б – випадок передавання символів 0 і 1).

Прийом ЧМн сигналу зазвичай здійснюється кореляційним методом, тобто обчисленням взаємної кореляції прийнятого сигналу з коливаннями-зразками (опорними сигналами), які є гармонічними коливаннями з використовуваними для маніпуляції частотами. На виході обирається той символ, частота якого виявляється максимально корельованою з вхідним сигналом.

При *фазовій маніпуляції* (ФМн, phase shift keying, PSK) стрибкоподібно змінюється фаза коливання-носія. Графік сигналу з двохпозиційною фазовою маніпуляцією представлено на рис. 6.5, в. На практиці фазова маніпуляція використовується при невеликому числі можливих значень початкової фази – як правило - 2, 4 або 8.

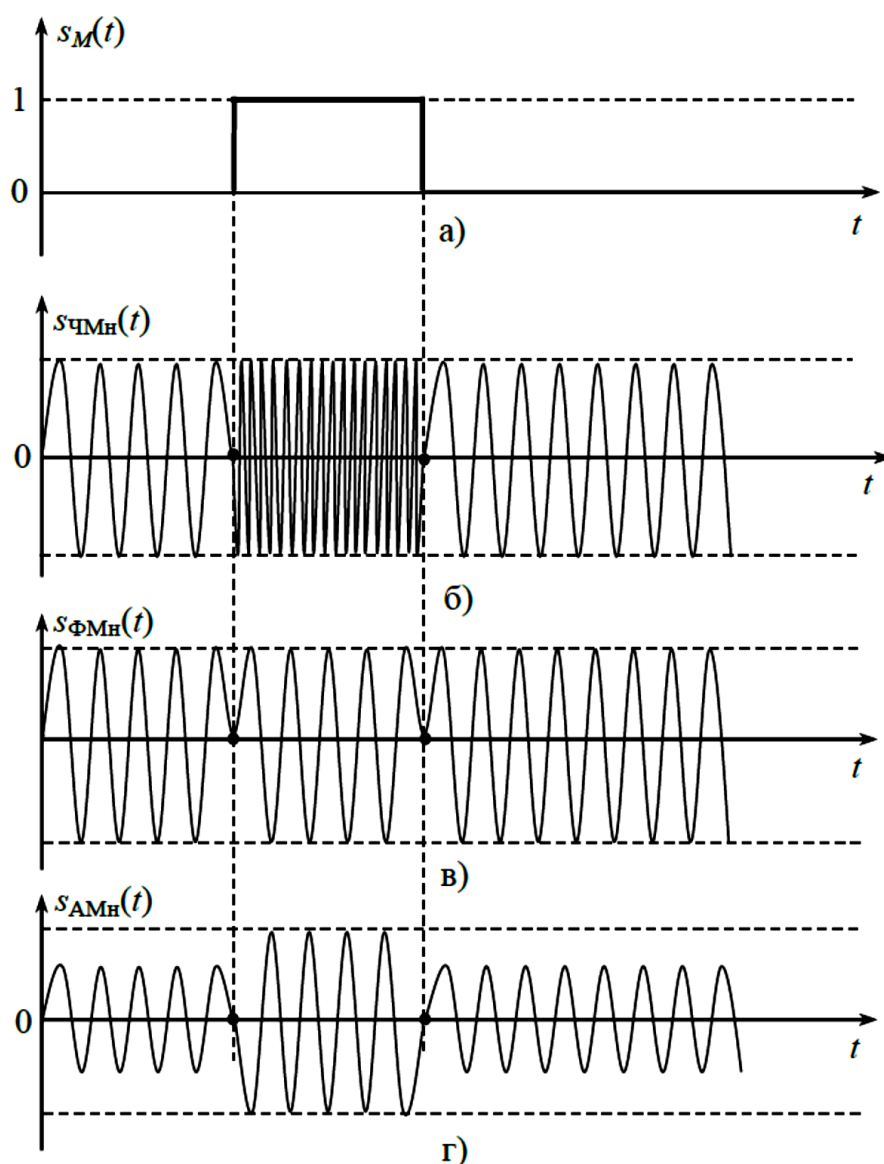


Рис. 6.5. Приклади маніпульованих сигналів: а) модулюючий бінарний сигнал; б) частотно-маніпульований; в) фазо-маніпульований; г) амплітудно-маніпульований

Крім того, при прийомі сигналу складно виміряти абсолютне значення початкової фази; значно простіше визначити відносний фазовий зсув між двома сусідніми символами. Тому зазвичай використовується

фазорізницева маніпуляція (синоніми – диференціальна фазова маніпуляція, відносна фазова маніпуляція, differential phase shift keying, DPSK).

При *амплітудній маніпуляції* (АМн; amplitude shift keying, ASK), стрибкоподібно змінюється амплітуда коливання-носія. Приклад АМн сигналу наведено на рис. 6.5, г.

6.2. Демодуляція сигналів

Для виділення на приймальному боці з високочастотного модульованого коливання корисного (модулюючого) сигналу потрібно виконати обернену до модулювання операцію демодуляції (детектування).

Детектори зазвичай будуються шляхом об'єднання нелінійних елементів з відповідними лінійними частотно-селективними колами.

Принцип детектування АМ сигналів розглянемо на прикладі ідеального детектора АМ коливання. Якщо на його вході діє АМ коливання

$$u_{\text{вх}}(t) = U_{\text{мвх}} (1 + M \cos \Omega t) \cos \omega_0 t,$$

то на виході буде отримано пропорційний переданому повідомленню низькочастотний сигнал $u_{\text{вх}}(t) = U_{\text{мвх}} \cos \Omega t$. Ефективність роботи детектора прийнято оцінювати коефіцієнтом детектування

$$k_{\text{дет}} = U_{\text{мвих}} / MU_{\text{мвх}}, \quad (6.7)$$

який є відношенням амплітуди низькочастотного сигналу на виході до «розмаху» зміни амплітуди високочастотного сигналу на вході.

Детектування може бути виконане, якщо подати АМ сигнал на безінерційний нелінійний елемент і здійснити подальшу фільтрацію низькочастотних складових спектра.

Фазове детектування виконується з допомогою фазових детекторів – пристроїв для демодуляції коливань з повною фазою $\Psi(t) = \omega_0 t + \varphi(t)$, яка є модульованою по фазовому куту. Фазовий детектор виробляє напругу пропорційну різниці фаз вхідних коливань, одне з яких не модульоване є *опорним* (формується допоміжним зовнішнім джерелом), а інше модульоване – *сигнальним*.

Частотне детектування. При частотній модуляції, як відомо, корисне повідомлення є пропорційним відхиленню миттєвої частоти сигналу від частоти-носія. Пристрій, напруга на виході якого пропорційна відхиленню

частоти від деякого опорного значення, називається частотним детектором.

Частотну модуляцію можна перетворити в амплітудну або фазову модуляцію з допомогою лінійного частотного фільтру з подальшим амплітудним або фазовим детектуванням. Також використовуються частотні детектори лічильного типу, в яких вихідна напруга є функцією кількості імпульсів, що надходять на вхід в одиницю часу.

Демодуляція сигналу з квадратурною модуляцією. Як й інші різновиди сигналів з КМ, квадратурно-модульований сигнал може демодулюватись шляхом множення на опорне колювання. Проте оскільки КАМ сигнал є сумою двох АМ сигналів, то і опорних колювань повинно бути два – із зсувом фаз на 90° .

Аналогічно демодулюється сигнал з квадратурною маніпуляцією. Як і у випадку аналогової квадратурної модуляції, сигнал множиться на два колювання-носія, що зсунуті по фазі одне відносно одного на 90° , а результати множення пропускаються через ФНЧ (рис. 6.6).

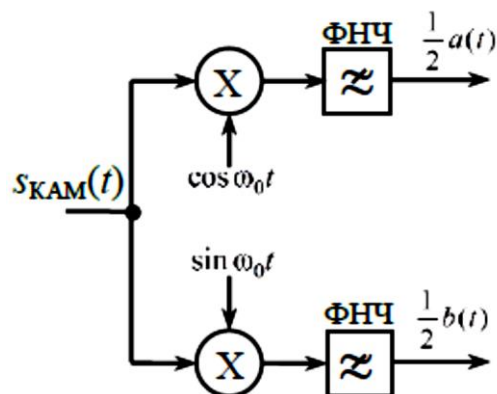


Рис. 6.6. Демодуляція КАМ-сигналу

На виходах цих ФНЧ будуть отримані аналогові сигнали синфазної і квадратурної складових. Далі ці сигнали дискретизуються з частотою, яка дорівнює символній швидкості. Пари відліків синфазної і квадратурної складових утворюють комплексне число, і відповідний цьому числу інформаційний символ видається як вихідний результат.

Питання для самоперевірки

1. Для чого виконують модуляцію сигналів?
2. Наведіть характеристику видів модуляції сигналів.

3. В чому полягає кутова модуляція сигналів?
4. Що таке «маніпуляція» сигналу?
5. В чому полягає односмугова модуляція?
6. Взаємозв'язок фазової і частотної модуляцій.
7. Демодуляція сигналів.
8. В чому сенс фазового і частотного детектування?

РОЗДІЛ 7. ОСОБЛИВОСТІ ЦИФРОВОЇ ОБРОБКИ ЗОБРАЖЕНЬ

Плоске монохромне зображення можна визначити як функцію (двовимірний розподіл) яскравості, оптичної щільності, рівня сірого тощо за координатами на площині. Кольорові зображення можуть формуватись комбінацією або накладанням кількох монохромних зображень. Якщо величини функції та координат приймають кінцеве число дискретних значень, то говорять про *цифрове зображення*. Цифрове зображення складається з кінцевого числа елементів, кожен з яких розташований у конкретному місці і приймає певне значення і називається елементом зображення або пікселем (слово pixel утворене від англійського словосполучення picture element - елемент зображення). Цифровою обробкою зображень називається обробка цифрових зображень за допомогою комп'ютерів.

7.1 Основи цифрової обробки зображень.

Не існує загальноприйнятої точки зору, де закінчується обробка зображень і починаються інші суміжні області, наприклад, аналіз зображень і машинний зір. Іноді розмежування робиться тут за принципом, що обробка зображень визначається як дисципліна, в якій на вході і на виході процесу присутні зображення. Такий поділ є обмеженим і штучним. Наприклад, у рамках цього визначення навіть така тривіальна задача, як визначення середньої інтенсивності по полю зображення, не може розглядатися як операція обробки зображення [12].

З іншого боку, існують такі галузі, як машинний зір, де кінцевою метою є комп'ютерна імітація людського зору, включаючи навчання, здатність до висновків та дій на основі спостережуваної інформації. Ця область сама собою утворює лише одне із напрямів штучного інтелекту, метою якого є імітація інтелектуальної діяльності.

Область, пов'язана з аналізом зображень (інакше звана «інтерпретація» зображень), займає проміжне положення між обробкою зображень і машинним зором.

У всьому діапазоні від обробки зображень до машинного зору немає чітких меж, проте можна розрізнити в ньому комп'ютеризовані процеси низького, середнього та високого рівня. Процеси низького рівня стосуються лише примітивних операцій типу попередньої обробки з метою зменшення шуму, підвищення контрасту або поліпшення різкості зображень.

Для низькорівневих процесів характерним є той факт, що на вході та на виході присутні зображення. Обробка зображень на середньому рівні охоплює такі завдання, як сегментація (розподіл зображення на області або виділення на ньому об'єктів), опис об'єктів і стиснення їх у зручну для комп'ютерної обробки форму, а також класифікація (розпізнавання) окремих об'єктів. Для процесів середнього рівня характерною є наявність зображень тільки на вході. На вихід надходять ознаки і атрибути, що витягуються з цих зображень (наприклад межі областей, лінії контурів, відмітні ознаки конкретних об'єктів). Нарешті, високорівнева обробка включає «осмислення» набору розпізнаних об'єктів, як це робиться в аналізі зображень, і в межі здійснення пізнавальних функцій, які прийнято пов'язувати із зором.

За призначенням обробку зображень можна розділити на таку, що використовується для відтворення зображення в полі зору оператора, і таку, що використовується у прикладних і наукових цілях. Задачі і методи обробки при цьому відрізняються, але частково можуть співпадати. В першому випадку основним завданням обробки є формування найбільш придатного для візуального сприйняття людиною зображення. В другому випадку важливим є пошук і виділення/підсилення певної інформації в зображенні, на основі якої автоматична система або людина-оператор можуть оцінити стан досліджуваного простору об'єктів.

Типові завдання 1-го напрямку: геометричні перетворення (обертання, масштабування); колірна корекція (зміна яскравості та контрасту, квантування кольору, перетворення на інший колірний простір тощо); комбінування зображень у різний спосіб; інтерполяція та згладжування; редагування та ретушування; розширення динамічного діапазону; компенсація втрати різкості.

Типові завдання 2-го напрямку: поділ зображення області; порівняння двох та більше зображень; розпізнавання тексту; стиснення зображень; обробка супутникових знімків; машинний зір; обробка даних для виділення релевантних ознак (виявлення об'єктів, ідентифікація особи, стеження за дорожньою розміткою, аналіз медичних знімків тощо).

Найбільш звичні для людського сприйняття є зображення, сформовані електромагнітним випромінюванням (від рентгенівського до інфрачервоного діапазону). Іншими джерелами, які можуть створювати зображення, є акустичні коливання, електронні пучки, що застосовуються в мікроскопії. Крім того, є цілий клас синтетичних (штучних) зображень, які синтезуються комп'ютерними програмами та використовуються для моделювання та візуалізації.

Основними стадіями цифрової обробки зображень є:

1. *Реєстрація зображення*. Реєстрація зображення може бути простою, як у випадку, коли вихідне зображення вже представлено в цифровій формі. Реєстрація зображення часто включає деяку попередню обробку, наприклад масштабування.

2. *Покращення зображення* - це процес маніпулювання зображенням, у результаті якого воно стає більш придатним конкретного застосування, ніж оригінал. Тут важливим є слово «конкретного», оскільки воно з самого початку встановлює, що методи поліпшення зображень є проблемно-орієнтованими. Так, наприклад, метод, який корисний для поліпшення рентгенівських зображень, може виявитися не кращим підходом для поліпшення супутникових зображень, знятих в інфрачервоному діапазоні електромагнітного спектру.

Загальної «теорії» покращення зображень не існує. Якщо зображення обробляється з метою візуальної інтерпретації, оцінку, наскільки добре працює конкретний метод, дає зрештою спостерігач.

3. *Відновлення зображень* - це область, також пов'язана з підвищенням візуальної якості зображення, однак, на відміну від власне покращення, критерії якого є суб'єктивними, відновлення зображення є об'єктивним у тому сенсі, що методи відновлення зображень спираються на математичні чи ймовірнісні моделі спотворень зображення.

Навпаки, поліпшення зображень ґрунтується на суб'єктивних уподобаннях людського сприйняття, які пов'язані з тим, що саме вважається «хорошим» результатом поліпшення.

4. *Обробка кольорових зображень* набула особливої важливості через значне розширення використання кольорових зображень в Інтернеті.

5. *Стиснення* відноситься до методів зменшення обсягу пам'яті, необхідного для зберігання зображення, або звуження смуги пропускання каналу, необхідної для передачі. Хоча техніка запам'ятовуючих пристроїв за останнє десятиліття була значно вдосконалена, цього не можна сказати щодо пропускної спроможності ліній зв'язку.

6. *Морфологічна обробка* пов'язана з інструментами для вилучення таких компонентів зображення, які можуть бути корисними для представлення та опису форми.

7. *Сегментація* поділяє зображення на складові або об'єкти. Наразі автоматична сегментація належить до найважчих завдань цифрової обробки зображень. Надмірно докладна сегментація спрямовує процес на довгий шлях вирішення завдання, вимагаючи ідентифікації об'єктів окремо. З іншого боку, недостатньо докладна або неправильна сегментація майже неминуче призведе до виникнення помилок на фінальній стадії обробки. Загалом, що точнішою є сегментація, то більшою є ймовірність правильного розпізнавання.

8. *Подання та опис* майже завжди слідує безпосередньо за етапом сегментації, на виході якого зазвичай є лише необроблені дані про пікселі, які або утворюють границі області (тобто дається безліч пікселів, що відокремлюють одну область зображення від іншої), або репрезентують всі точки самих областей. В обох випадках необхідно перетворити дані на форму, придатну для комп'ютерної обробки.

Перше рішення, яке слід ухвалити, - чи мають ці дані подаватися у формі лише меж областей чи областей цілком. Подання у вигляді границь підходить для тих випадків, коли в центрі уваги знаходяться зовнішні характеристики форми областей, наприклад кути та вигини.

Подання у вигляді областей є кращим, якщо акцент робиться на внутрішніх властивостях об'єктів, наприклад, текстурі або формі скелета. У

деяких додатках ці уявлення доповнюють одне одного. Вибір способу подання - лише частина прийняття рішення щодо перетворення «сирих» піксельних даних у відповідну для подальшої комп'ютерної обробки форму.

Повинен бути ще зазначений метод опису даних, при якому висувалися б на передній план релевантні ознаки. Побудова опису (вибір ознак) пов'язане з виділенням атрибутів, які б виражали корисну інформацію в кількісному вигляді або могли бути основою для розрізнення класів об'єктів.

9. *Розпізнавання* є процесом, який присвоює деякому об'єкту ідентифікатор (наприклад, «транспортний засіб») на підставі його описників.

Первинне зображення має неперервні координати x та y , а також неперервну амплітуду A . Переведення його в цифрову форму вимагає перетворення координат і амплітуди до дискретних значень. Зображення координат скінченною кількістю значень називається *дискретизацією*, а зображення амплітуди значеннями зі скінченного набору відліків називається *квантуванням*. Якщо координати x і y , а також амплітуда A вибираються з скінченних наборів величин, то зображення називається цифровим. З математичної точки зору результатом дискретизації монохромного зображення є матриця чисел.

В загальному випадку, частота дискретизації може бути різною по різним координатам (x, y) : f_{Dx}, f_{Dy} (з відповідними кроками дискретизації T_{Dx}, T_{Dy}). Як правило, дискретизація зображень проводиться не в часовій області, а в просторовій, і частота дискретизації показує, скільки відліків зображення буде виміряне на одиницю довжини по кожній координаті. Розмірність частоти дискретизації в цьому випадку буде $f_{Dx} \cdot f_{Dy}$, а крок дискретизації буде виражений в одиницях довжини.

Дискретне зображення задане своїми значеннями $I[n, m]$ в деяких точках площини. В реальних застосуваннях дискретне зображення містить скінченну кількість елементів, які називаються пікселями (pixel, від англійського picture element – елемент зображення). В цьому випадку дискретна функція $I[n, m]$ має лише скінченну кількість значень:

$$I_1[n, m], n = \overline{0, N-1}, m = \overline{0, M-1}.$$

Таке зображення можна представити у вигляді матриці, кожний елемент якої дорівнює значенню яскравості відповідного пікселя:

$$I_1[n, m] = \begin{bmatrix} I_1[0,0] & I_1[0,1] & \dots & I_1[0, M-1] \\ I_1[1,0] & I_1[1,1] & \dots & I_1[1, M-1] \\ \vdots & & \ddots & \vdots \\ I_1[N-1,0] & I_1[N-1,1] & & I_1[N-1, M-1] \end{bmatrix}$$

Нумерація пікселів є подвійною - перший індекс відповідає номеру рядка, другий – номеру стовпця. Лівий верхній піксель зображення має номер $[0, 0]$.

Всі методи аналізу та обробки зображень зводяться до математичних операцій над неперервною функцією $A(x, y)$, яка описує неперервне зображення, або над пікселями $I[n, m]$ (у випадку дискретного зображення).

Якщо значення яскравості кожного пікселя може приймати лише значення 0 або 1, то таке зображення називається бінарним, і пікселі можуть бути тільки двох кольорів (як правило, чорного та білого). У випадку, коли $I_1[n, m]$ може набувати будь-яке значення від 0 до 1, то говорять, що таке зображення є напівтоновим, побудоване в градаціях сірого (grayscale image).

Зображення не завжди мають цифрову форму. Тому їх перетворення у такий вигляд є обов'язковою операцією, якщо передбачається використовувати цифрову обробку, передавання, зберігання.

Як і у випадку одновимірних сигналів, це перетворення поєднує дві процедури: дискретизацію, яка полягає у заміні неперервного кадру дискретним, і квантування, яка виконує заміну неперервної множини значень яскравості множиною квантованих значень. За цифрового представлення кожному із квантованих значень яскравості ставиться у відповідність двійкове число.

Заміну неперервного зображення дискретним можна виконати різними способами. Можна, наприклад, вибрати яку-небудь систему ортогональних функцій і, вирахувавши коефіцієнти представлення зображення за цією схемою (за цим базисом), замінити ним зображення. Різноманітність базисів уможливорює утворення різних дискретних представлень неперервного зображення. Однак найбільш вживаною є періодична дискретизація, а саме дискретизація з прямокутним растром.

Такий спосіб дискретизації може розглядатись як один із варіантів використання ортогонального базису, у якому використовуються як елементи зсунуті δ -функції. За умови використання прямокутного растра в остаточному вигляді сигнали оцифрованого зображення подають у вигляді матриці (монохромне зображення) або набору матриць (кольорове зображення), рядки та стовпці яких містять квантовані значення параметрів відповідних елементів дискретизованого зображення.

Основне питання, яке виникає під час заміни неперервного зображення дискретним, полягає у визначенні умов, за яких така заміна є повноцінною, тобто не супроводжується втратою інформації, що міститься в неперервному сигналі. Втрати відсутні, якщо, маючи в розпорядженні дискретний сигнал, можна відновити неперервний. З математичного погляду питання полягає у відновленні неперервного сигналу в двовимірних проміжках між вузлами, у яких його значення відомі, тобто в здійсненні двовимірної інтерполяції.

Під час цифрової обробки зображень динамічний діапазон значень яскравості $[dL, dL+1]$ (припускається, що він обмежений) ділять на інтервали. Якщо яскравість x відліку належить інтервалу (d_j, d_j+1) , тобто $(d_j < x \leq d_j+1)$ вихідний відлік замінюється на рівень квантування r_j , де $d_j; j=1, L+1$ - пороги квантування. Задача побудови квантувача полягає у визначенні значень порогів d_j та рівнів r_j . Найпростіший спосіб розв'язання цієї задачі полягає у розбитті динамічного діапазону на однакові інтервали. Однак таке рішення не буде найкращим, оскільки, у разі наявності групування значень яскравості відліків зображення в певних областях доцільно квантувати нерівномірно. Це дасть змогу зменшити помилку квантування.

Отже, задачу побудови квантувача можна сформулювати як задачу знаходження оптимальних значень d_j та r_j , що задовольняли б певний критерій оптимізації. Як правило, за фіксованої кількості рівнів квантувач оптимізуються за критерієм мінімальної середньоквадратичної похибки.

У системах цифрової обробки зображень прагнуть зменшити кількість рівнів та порогів квантування, оскільки від їхньої кількості залежить довжина двійкового кодового слова, яким представляються проквантовані

відліки в комп'ютері. Однак за порівняно невеликої кількості рівнів L на проквантованому зображенні з'являються хибні контури, які виникають внаслідок стрибкоподібної зміни яскравості проквантованого зображення і особливо помітні на пологих ділянках її зміни. Вони значно погіршують візуальну якість зображення, оскільки зір людини особливо чутливий якраз до контурів.

Щоб уникнути нерівномірного квантування, яке не можна реалізувати за допомогою стандартного АЦП, використовують нелінійне перетворення. Здійснюють нелінійне перетворення відліку вихідного зображення, щоб щільність розподілу ймовірностей перетворених відліків була рівномірною, тобто виконується процедура еквалізації. Потім відліки квантують з рівномірним кроком та виконують зворотне нелінійне перетворення.

Для руйнування хибних контурів перед рівномірним квантуванням до відліків яскравості додають шум з рівномірною щільністю розподілу ймовірностей, який переводить одні відліки зображення на рівень вище, а інші - на рівень нижче. Дисперсія шуму, який додається, повинна бути невеликою, щоб не призвести до спотворень, які сприймаються як "сніг" на зображенні, і водночас достатньою для руйнування хибних контурів.

Визначення реального часу залежить від конкретної задачі і пов'язане з обсягом обчислень алгоритму, точністю обчислень та частотою (періодом) дискретизації. Цифрова система працюватиме в режимі реального часу, якщо час виконання алгоритму не перевищуватиме періоду дискретизації. Це означає, що залишається іще деякий запас часу, який називають часом очікування.

7.2. Цифрова обробка кольорових зображень

Характеристики кольору

Чорно-білі зображення, які описуються розподілом яскравості у площині, є частковим випадком ширшого класу кольорових зображень. Колір є суб'єктивним поняттям та залежить від особливостей зорового сприйняття. Колір можна характеризувати трьома величинами: світлотою, колірним тоном і насиченістю.

Світлота відіграє провідну роль у сприйнятті зображень. Одні й ті ж

кольори, що мають різну світлоту, викликають різні зорові відчуття. Наприклад, білий колір залежно від світлоти може сприйматися як білий, світло-сірий, сірий, темно-сірий тощо. Залежно від яскравості сусідніх ділянок зображення нейтральний колір, що характеризується однією і тією ж яскравістю, може сприйматися і як білий, і як чорний (яскравість чорних літер газетного тексту на сонці вище, ніж яскравість білого паперу, на якому вони надруковані, в тіні). Тобто світлота, що є зоровим відчуттям, є нелінійною і досить складною функцією яскравості, яка визначається властивостями зору. Яскравість є фізичною величиною і від властивостей зору не залежить.

Колірний тон випромінювання характеризує властивість кольору, що відрізняє його від білого та сірого. Назви кольорів (червоний, жовтий, синій тощо) можна розглядати як приблизні позначення колірному тону. Кількісно колірний тон випромінювання характеризується домінуючою довжиною хвилі, тобто довжиною хвилі такого спектрально чистого кольору, який необхідно змішати з білим кольором у певній пропорції для того, щоб отримати описуваний колір. Між колірним тоном та домінуючою довжиною хвилі існує нелінійна залежність. Наприклад, жовтий колір займає на шкалі довжин хвиль інтервал приблизно 30 нм, тоді як червоний - близько 75 нм.

Насиченість кольору характеризує ступінь розбавленості спектрально чистого кольору білим. Якщо до спектрально чистого кольору додати білий, то насиченість зменшиться. Частка спектрального кольору у його суміші з білим називається чистотою кольору p

$$p = \frac{L_{\lambda}}{L_{\lambda} + L_E} = \frac{L_{\lambda}}{L},$$

де L_{λ} - яскравість спектрально чистого кольору; L_E - яскравість рівноенергетичного білого кольору (тобто випромінювання, спектральна інтенсивність якого на всіх довжинах хвиль видимого випромінювання є сталою); L - яскравість суміші.

Вочевидь, що з спектрально чистих випромінювань чистота кольору є максимальною ($p = 1$), а рівноенергетичного білого - мінімальною ($p = 0$). З огляду на особливості зорової системи людини зв'язок між насиченістю

кольору (тобто зоровим відчуттям насиченості) і чистотою кольору є нелінійною. Більше того, при одній і тій же чистоті кольору, червоний і синій спектрально чисті кольори сприймаються як високонасичені, а жовтий і блакитний (теж спектрально чисті) - менш насичені кольори. Таким чином, між характеристиками сприйняття кольору: світлою, кольоровим тоном і насиченістю, з одного боку, і характеристиками кольору: яскравістю, домінуючої довжиною хвилі та чистотою кольору, з іншого боку, мають місце нелінійні залежності.

Колірні розрахунки (колориметрія) базуються на так званих аксіомах зрівнювання (законах змішування) кольорів, сформульованих Грасманом в 1854 р. Фізіологічна основа колориметрії полягає в наявності в сітківці ока трьох типів світлочутливих клітин - колбочок, які різняться між собою спектральною чутливістю.

Аксіоми зрівнювання кольорів

1. Перший закон змішування кольорів: будь-який колір може бути отримано шляхом змішування не більше трьох лінійно незалежних кольорів, тобто таких кольорів, коли жоден з них не є результатом змішування двох інших.

2. Колір, отриманий за однієї інтенсивності світла, зберігається у широкому діапазоні зміни інтенсивностей.

3. Третій закон змішування кольорів: колір суміші декількох випромінювань визначається тільки кольорами компонентів, що змішуються, незалежно від того, як вони отримані. Тобто суміш кольорів не може бути розділена зоровою системою людини на окремі компоненти.

4. Яскравість суміші кольорів дорівнює сумі яскравостей її компонентів.

5. Закон складання. Якщо колір M' еквівалентний кольору N' , а колір P' - кольору Q' , (тобто якщо для зорової системи вони не відрізняються) то суміш кольорів M' та P' є еквівалентною суміші кольорів N' та Q' . Записується це твердження так: якщо $M' = N'$ і $P' = Q'$, то $M' + P' = N' + Q'$.

(Тут і далі в аналогічних висловлюваннях знак "=" означає еквівалентність).

6. Закон віднімання. Якщо суміш кольорів M' і P' еквівалентна суміші

кольорів N' та Q' , а колір P' еквівалентний кольору Q' , то кольори M' та N' також еквівалентні один одному. Це твердження може бути записано таким чином: якщо $M' + P' = N' + Q'$, а $P' = Q'$, то $M' = N'$.

7. Закон транзитивності. Відповідно до цього закону: якщо колір M' є еквівалентним кольору N' , а колір N' еквівалентний кольору P' , то колір M' є еквівалентним кольору P' .

8. Зрівнювання кольорів: для будь-якого кольору C' справедливе одне із трьох співвідношень:

- колір C' урівнюється сумішшю $A_1(C)$ частки одиничної кількості кольору M , $A_2(C)$ частки одиничної кількості кольору N та $A_3(C)$ частки одиничної кількості кольору P ;

- суміш кольору C' з $A_1(C)$ часткою одиничної кількості кольору M урівнюється сумішшю $A_2(C)$ частки одиничної кількості кольору N та $A_3(C)$ частки одиничної кількості кольору P ;

- суміш кольору C' з $A_1(C)$ часткою одиничної кількості кольору M та $A_2(C)$ часткою одиничної кількості кольору N дорівнює $A_3(C)$ частки одиничної кількості кольору P .

Яке з трьох співвідношень буде виконано, залежить від кольорів, що змішуються.

Аксіоми Грасмана доповнюють ще одним експериментально встановленим законом, який стверджує, що безперервній зміні фізичних характеристик випромінювання відповідає безперервна зміна кольору.

Принципи побудови колориметричних систем

З восьмої аксіоми Грасмана випливає, що колір C' може бути отриманий як суміш трьох основних кольорів P_1 , P_2 та P_3 , тобто

$$C' = A_1(C)P_1 + A_2(C)P_2 + A_3(C)P_3, \quad (7.1)$$

де $A_1(C)$, $A_2(C)$, $A_3(C)$ - інтенсивності основних кольорів, які можуть набувати як позитивних, так і негативних значень.

Оскільки інтенсивності некогерентних джерел світла додаються лінійно, то спектральна густина суміші дорівнює сумі спектральних густин її компонентів.

З формули (7.1) видно, що спектральна густина $C(\lambda)$ може бути замінена еквівалентною спектральною густиною суміші основних кольорів

$$C(\lambda) = A_1(C)P_1(\lambda) + A_2(C)P_2(\lambda) + A_3(C)P_3(\lambda), \quad (7.2)$$

де λ - довжина світлової хвилі.

Тобто, випромінювання зі спектральною густиною $C(\lambda)$ і випромінювання зі спектральною густиною $A_1(C)P_1(\lambda) + A_2(C)P_2(\lambda) + A_3(C)P_3(\lambda)$ викликають однакові відчуття кольору.

Виражаючи інтенсивність основних кольорів у формулі (7.2) через координати кольору відповідно до формули (7.1), отримуємо

$$C(\lambda) = T_1(C)A_1(W)P_1(\lambda) + T_2(C)A_2(W)P_2(\lambda) + T_3(C)A_3(W)P_3(\lambda). \quad (7.3)$$

Із четвертої аксіоми Грасмана витікає, що яскравість кольору $L(C)$ дорівнює сумі яскравостей його компонентів

$$L(C) = \int_{\lambda_{\min}}^{\lambda_{\max}} C(\lambda)V(\lambda)d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} [A_1(C)P_1(\lambda) + A_2(C)P_2(\lambda) + A_3(C)P_3(\lambda)]V(\lambda)d\lambda,$$

де $V(\lambda)$ – функція відносної спектральної світлової ефективності монохроматичного випромінювання для денного зору (крива видимості ока); λ_{\min} , λ_{\max} – граничні довжини хвиль електромагнітного випромінювання видимого діапазону спектру.

Подаючи інтенсивність основних кольорів через координати кольору, знаходимо

$$L(C) = T_1(C)L(P_1) + T_2(C)L(P_2) + T_3(C)L(P_3), \quad (7.4)$$

де

$$L(P_1) = \int_{\lambda_{\min}}^{\lambda_{\max}} A_1(W)P_1(\lambda)V(\lambda)d\lambda,$$

$$L(P_2) = \int_{\lambda_{\min}}^{\lambda_{\max}} A_2(W)P_2(\lambda)V(\lambda)d\lambda,$$

$$L(P_3) = \int_{\lambda_{\min}}^{\lambda_{\max}} A_3(W)P_3(\lambda)V(\lambda)d\lambda \quad - \text{ коефіцієнти яскравості основних}$$

кольорів.

Формули (7.3) та (7.4) є кількісною основою колориметрії. Для створення колориметричної системи необхідно:

1. Вибрати основні кольори та визначити одиниці їх вимірювання P_1 , P_2 , P_3 . Оскільки вибір основних кольорів нічим не обмежений, то й кількість колориметричних систем, які можна створити, також є не обмеженою.

2. Визначити для обраних основних кольорів координати кольору спектральних випромінювань одиничної потужності, які називаються питомими координатами чи кривими змішування. Знання питомих координат дозволяє визначати координати кольору випромінювання з довільною спектральною густиною $C(\lambda)$ шляхом зваженого підсумовування $\overline{T_1(\lambda)}$, $\overline{T_2(\lambda)}$, $\overline{T_3(\lambda)}$. При цьому функцією, що зважує, є спектральна густина

$$\begin{aligned}
 T_1(C) &= \int_{\lambda_{\min}}^{\lambda_{\max}} C(\lambda) \overline{T_1(\lambda)} d\lambda, \\
 T_2(C) &= \int_{\lambda_{\min}}^{\lambda_{\max}} C(\lambda) \overline{T_2(\lambda)} d\lambda, \\
 T_3(C) &= \int_{\lambda_{\min}}^{\lambda_{\max}} C(\lambda) \overline{T_3(\lambda)} d\lambda.
 \end{aligned}
 \tag{7.5}$$

З формул (7.5) випливає, що будь-який колір може бути повністю описаний, якщо задати три координати $T_1(C)$, $T_2(C)$, $T_3(C)$. Тобто колір може бути представлений положенням точки в тривимірному просторі T_1 , T_2 , T_3 , або вектором (рис. 7.1).

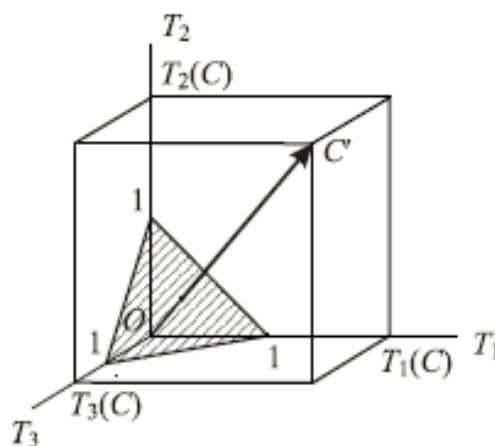


Рис. 7.1. Графічне подання кольору в колірному просторі

При цьому деякі координати кольору можуть набувати від'ємних значень, проте результуюча яскравість, яка визначається формулою (7.4), завжди буде додатною. Від'ємні значення координат кольору інтерпретуються наступним чином: основний колір, що відповідає негативному значенню координати, має бути доданий до вимірюваного

кольору для досягнення зрівнювання сумішшю двох основних кольорів, що залишилися. Це зауваження має принциповий характер. Так, наприклад, колір, у якого хоча б одна координата є від'ємною, не може бути отриманий шляхом фізичного змішування трьох основних кольорів.

Часто при описі кольору його яскравість не має значення. Тоді від координат кольору переходять до координат колірності (координат одиничних кольорів) $t_1 = \frac{T_1}{T_1 + T_2 + T_3}$, $t_2 = \frac{T_2}{T_1 + T_2 + T_3}$, $t_3 = \frac{T_3}{T_1 + T_2 + T_3}$, які повністю визначають колірний тон та насиченість кольору.

Співвідношення $t_1 + t_2 + t_3 = 1$ визначає площину одиничних відтінків, показану на рис. 7.1 штрихуванням, а координати одиничних кольорів - точку «проколу» цієї площини вектором кольору. Оскільки $t_3 = 1 - t_1 - t_2$, то для повного опису колірності достатньо використовувати лише дві змінні t_1 і t_2 , що зручно при поданні результатів колірних розрахунків на двовимірному графіку. Перехід до такого графіка досягається шляхом проєкції площини одиничних кольорів на площину T_1OT_2 . При цьому рівносторонній трикутник одиничних кольорів, показаний на рис. 7.1 штрихуванням, перетворюється на прямокутний трикутник.

Розглянемо колориметричні системи, що найчастіше застосовуються на практиці.

Колориметрична система RGB

У 1931 р. міжнародною комісією з освітлення була розроблена та рекомендована до застосування колориметрична система RGB.

Особливістю цієї колориметричної системи є використання як основних кольорів трьох монохроматичних випромінювань з довжинами хвиль: для червоного (R) – 700 нм, для зеленого (G) – 546,1 нм та для синього (B) – 435,8 нм.

Одиниці вимірювання координат кольору, які в цій системі позначають через r' , g' , b' обрано таким чином, щоб для опорного білого кольору E , що характеризується постійною спектральною інтенсивністю у видимій частині спектра, вони були б однакові, тобто $r' = g' = b'$.

Використання як основних реальних кольорів R, G, B є зручним, оскільки дозволяє при колориметричних вимірюваннях застосовувати

метод зрівнювання кольорів. При цьому колір C' , що вимірюється, може бути записаний як $C' = r'R + g'G + b'B$, а його яскравість як $L = r' + 4,5907g' + 0,0601b'$, де R, G та B - одиниці червоного, зеленого та синього відповідно.

З наведеної формули випливає, що найбільший внесок у яскравість робить зелена компонента кольору, а найменший - синя компонента.

Колориметрична система XYZ

Одночасно з колориметричною системою RGB міжнародною комісією з освітлення в 1931 р. було розроблено та рекомендовано іншу колориметричну систему XYZ (CIE). Викликано це було тим, що поряд з перевагами системи RGB, основною з яких є зручність її застосування при вимірюванні кольору, мають місце два істотні недоліки, що роблять її незручною при колірних розрахунках. Перший полягає в наявності від'ємних коефіцієнтів, другий - у необхідності визначати яскравість кольору шляхом зваженого підсумовування трьох координат кольору.

Основні кольори системи XYZ отримали із перерахунку системи RGB. Їх вибрано так, що:

- координати кольору для будь-якого випромінювання завжди є додатними;
- яскравість випромінювання визначається однією координатою кольору;
- точка, що відповідає опорному білому кольору типу E (білому рівноенергетичному випромінюванню), знаходиться в центрі тяжіння колірної трикутника.

Основні кольори X, Y, Z не існують в природі, а є певними математичними описами. Колір у цій колориметричній системі визначається як

$$C' = x'X + y'Y + z'Z,$$

де x' , y' , z' - координати кольору.

Недоліком колориметричних систем RGB та XYZ є те, що в них однаковим відстаням на графіках колірності не відповідають однакові візуальні відмінності. Прагнення позбавитися цього недоліку призвело до розробки низки так званих рівноконтрастних систем, деякі з яких розглянуто нижче.

Колориметрична система Мак-Адама

Колориметрична система Мак-Адама є рівноконтрастною, в якій однаковим відстаням на графіку кольорів відповідають приблизно однакові візуальні відмінності. Це робить її зручною для колірних розрахунків. Систему побудували на базі системи XYZ шляхом нелінійного перетворення координат колірностей. Нові рівноконтрастні координати колірності u , v , що визначають колірність випромінювання, знаходять як:

$$u = 4x(-2x + 12y + 3),$$

$$v = 6y(-2x + 12y + 3).$$

*Колориметрична система $L^*a^*b^*$*

Корекція яскравості та кольору застосовуються у різних галузях обробки зображень. Оскільки практично результати перетворень оцінюються за зображенням на моніторі, то необхідно забезпечити високу ступінь колірної відповідності між моніторами і можливими пристроями виведення. Насправді монітор повинен точно відтворювати кольори вихідного зображення, представленого в цифровому форматі, так само як і остаточні кольори друкованих зображень. Найкращим чином це досягається при використанні незалежної від пристрою колірної моделі, яка пов'язує між собою колірні охоплення моніторів, пристроїв виведення, а також інших пристроїв, що використовуються. Прикладом такої моделі у багатьох системах керування кольором є $L^*a^*b^*$ (CIE LAB). Колориметрична система $L^*a^*b^*$ була рекомендована МКО у 1976 р. Ця система є рівноконтрастною. Її отримано із колориметричної системи XYZ шляхом перетворення її колірних координат за формулами

$$L^* = 116 \cdot h\left(\frac{Y}{Y_w}\right) - 16, \quad a^* = 500 \left[h\left(\frac{X}{X_w}\right) - h\left(\frac{Y}{Y_w}\right) \right], \quad b^* = 200 \left[h\left(\frac{Y}{Y_w}\right) - h\left(\frac{Z}{Z_w}\right) \right],$$

де

$$h(q) = \begin{cases} \sqrt[3]{q}, & \text{якщо } q > 0,008856, \\ 7,787q + 16/116, & \text{якщо } q \leq 0,008856, \end{cases}$$

а величини X_w , Y_w , Z_w є координатами опорного білого кольору.

Зазвичай фізичним джерелом такого опорного білого кольору є відбиваюча дифузна поверхня, яка освітлюється джерелом D65 стандарту МКО (на діаграмі колірності МКО йому відповідають координати $x=0,3127$,

$y=0,3290$).

Колірний простір $L^*a^*b^*$ є колориметричним (тобто кольори, що однаково сприймаються зоровою системою, мають однакові колірні координати). Також він є рівноконтрастним (тобто однаковим змінам координат колірності відповідають однакові зміни у відчутті кольору) і не залежить від пристрою відображення. Хоча цей колірний простір не може бути відображений безпосередньо (для подання на екрані або під час друку необхідний перехід в інший колірний простір), його колірне охоплення включає весь видимий спектр і дозволяє точно уявити кольори будь-яких моніторів, принтерів та інших пристроїв введення/виводу. Подібно до системи HSI, система $L^*a^*b^*$ дозволяє добре розрізняти інтенсивність (яка представлена яскравістю L^*) та колірність (яка представлена двома різницями кольорів: a^* - червоний мінус зелений і b^* - зелений мінус синій). Ця властивість робить систему $L^*a^*b^*$ дуже зручною як для поліпшення зображень (тональної і колірної корекції), так і для їх стиснення. В цій системі працюють багато промислових колориметрів. Вона використовується в основних графічних редакторах, таких як Photoshop, Corel PHOTO-PAINT, CorelDRAW. При використанні цієї системи в графічних редакторах значення координат кольору наводяться в умовних одиницях. Причому в різних редакторах вони можуть різнитися між собою. Так у графічному редакторі Corel PHOTO-PAINT $0 \leq L^* \leq 100$, $-60 \leq a^* \leq 60$, $-60 \leq b^* \leq 60$, а у графічному редакторі Photoshop $0 \leq L^* \leq 100$, $-128 \leq a^*$, $-128 \leq b^* \leq 127$. Для білого кольору $a^* = b^* = 0$.

Колориметричні системи HLS та HSB

Ці системи часто використовуються дизайнерами. Їх назви є скороченнями від слів: hue, lightness, saturation - колірний тон, світлота, насиченість, у першому випадку, і hue, saturation, brightness - колірний тон, насиченість, яскравість, у другому випадку. Значення колірного тону у цих системах задається у градусах, як і в колориметричній системі Манселла, а насиченість і світлота (яскравість у разі HSB) - у відсотках. Обидві колориметричні системи, маючи багато спільного, відрізняються у деталях. Так у системі HSB всі три координати кольору є взаємно незалежними, а в системі HSL координата L залежить від насиченості S. Взаємна

незалежність координат кольору створює певні зручності при обробці зображень, тому що дозволяє окремо коригувати спотворення яскравості та кольору.

Колориметричні системи CMY та CMYK

Ці колориметричні системи призначені для розрахунків кольорів при роботі з пристроями, що виводять зображення на папір. До таких пристроїв відносяться принтери та друкарські верстати. Назва колориметричної системи CMY утворена з початкових літер назв кольорів, прийнятих як основні в цій системі: Cyan - блакитний, Magenta - пурпуровий, Yellow - жовтий. Ці кольори є додатковими до основних кольорів, що використовуються у колориметричній системі RGB (червоному, зеленому, синьому).

У колориметричній системі CMYK додатково введено ще одну координату кольору К - ключовий чорний колір. Необхідність введення додаткової координати (надмірної, з погляду опису кольору) обумовлена прагненням розширити діапазон відтворюваних під час друку кольорів. Крім того, це дозволило покращити якість відтворення чорних та сірих тонів. Ці колориметричні системи представлені в сучасних графічних редакторах, призначених для підготовки до друку кольорових зображень.

Колориметрична система Манселла

В основу цієї системи покладено три характеристики кольору, що позначаються термінами: hue - колірний тон, chroma - насиченість та value - світлота. Систему реалізовано у вигляді атласу зразків з приблизно рівними щаблями зміни кольору, що охоплюють переважну частину області можливих кольорів. Ці зразки згруповані у просторове колірне тіло, причому ахроматичні кольори (від чорного до білого) розташовувалися по вертикальній осі, колірні тони - по колу, а насиченість - по радіусах. Наразі цю систему використовують для визначення кольору різних матеріалів. Серйозним її недоліком є складність перерахунку результатів вимірювань, отриманих у ній, до інших колориметричних систем.

Звісно, що для опису кольорового зображення можуть використовуватися й інші кольорові координатні системи, які при рішенні певних задач дозволяють досягти переваг у порівнянні з розглянутими вище

системами.

Обробка зображень може проводитися для: художнього перетворення зображення для досягнення певних ефектів; візуального (корекція яскравості та контрасту, корекція кольору тощо) або об'єктивного (корекція дисторсії, усунення змазу, розфокусування тощо) поліпшення якості зображення; аналізу зображення для вимірювань певних характеристик об'єктів, що досліджуються (аналіз інтерферограм, гартманогам, ФРТ тощо); розпізнавання образів (розпізнавання номерів автомобілів, відбитків пальців, пошук обличч тощо). Поставлені цілі досягають вирішенням наступних завдань: дискретизація, квантування та кодування зображень, геометричні перетворення, логічні та арифметичні операції над зображеннями, фільтрація, препарування зображень.

Деякі методи обробки напівтонових зображень застосовуються і до кольорових зображень. Інші методи мають бути модифіковані для узгодження з властивостями колірному простору.

Людина здатна розрізнати тисячі відтінків кольору і лише близько двох десятків відтінків сірого. Тому іноді напівтонове чорно-біле зображення перетворюють на псевдокольорове. Це реалізують шляхом співставлення кольорів з деякими значеннями сигналу інтенсивності монохромного зображення або деякими діапазонами зміни його інтенсивності. Таким чином, обробку кольорових зображень можна умовно розділити на обробку зображень у натуральних кольорах (коли зображення формуються пристроями реєстрації із спектральним розділенням) та обробку зображень у псевдокольорах. Перший напрямок є значно ширшим за другий і частково дублює його.

Підходи, що використовуються при обробці кольорових зображень, можна розділити на два напрямки: по-компонентну обробку та векторну обробку. Підходи першої категорії припускають, що кожна компонента кольору зображення обробляється окремо, а потім результуюче кольорове зображення складається з компонентів, оброблених окремо. Підходи другої категорії передбачають безпосередню роботу із кольоровими пікселями. Оскільки кольорове зображення містить як мінімум три складові частини, то значення кольорового пікселя є вектором. У RGB-моделі, наприклад,

значення пікселя зображення може розглядатися як вектор, проведений з початку координат у відповідну точку колірному простору. Те, що пікселі зображення є кольоровими, призводить до можливості обробляти незалежно кожну компоненту кольору зображення, використовуючи звичайні методи обробки напівтонових зображень. Однак результати, що отримуються таким способом, не завжди збігаються з результатами обробки, що виконується безпосередньо в колірному векторному просторі.

Перетворення кольорів полягає в обробці компонент кольорового зображення в рамках однієї окремо взятої колірної моделі. Це відрізняє такі перетворення від перетворень координат кольору під час переходу з однієї колірної моделі в іншу (наприклад, з моделі RGB модель HSI і назад).

Перетворення кольорових зображень можна записати як

$$g(x, y) = T[f(x, y)], \quad (7.6)$$

де $g(x, y)$ - перетворене (оброблене) кольорове зображення на виході, $f(x, y)$ - кольорове зображення на вході, T - оператор обробки, що діє на зображення f у просторовій околиці точки (x, y) .

Розглянемо колірні перетворення виду

$$s_i = T_i(r_1, r_2, \dots, r_n), i = 1, 2, \dots, n,$$

де змінні r_i і s_i - позначення колірних компонентів зображень $f(x, y)$ та $g(x, y)$ у довільній точці (x, y) , n - число колірних компонентів, $\{T_1, T_2, \dots, T_n\}$ - множина функцій перетворення або колірного відображення, які, діючи на величини r_i , дають величини s_i .

Вся сукупність n функцій перетворення T_i визначає єдине відображення T у виразі (7.6). Значення n визначається колірним простором, вибраним для опису пікселів зображень f та g . Наприклад, якщо використовується колірний простір RGB, то $n = 3$ та змінні r_1 , r_2 та r_3 позначають червону, зелену та синю компоненти вхідного зображення. Для колірних просторів CMYK та HSI - відповідно $n = 4$ та $n = 3$.

Теоретично у межах кожної з колірних моделей може бути здійснено будь-яке перетворення. Проте на практиці деякі операції краще пристосовані для реалізації у межах тієї чи іншої моделі. При ухваленні рішення щодо того, в якому колірному просторі реалізовувати це перетворення, необхідно враховувати витрати на перехід від одного

колірного подання до іншого. Наприклад, потрібно змінити інтенсивність кольорового зображення за допомогою перетворення

$$g(x,y) = kf(x,y), \text{ де } 0 < k < 1.$$

У колірному просторі HSI цього можна досягти за допомогою простого перетворення

$$s_3 = kr_3, \quad (7.7)$$

причому $s_1 = r_1$ і $s_2 = r_2$ і зміні підлягає лише компонент інтенсивності.

У колірному просторі RGB всі три компоненти мають бути зміненими:

$$s_i = kr_i, \quad i = 1, 2, 3. \quad (7.8)$$

Простір CMY вимагає застосування схожого набору лінійних перетворень

$$s_i = kr_i + (1-k), \quad i = 1, 2, 3. \quad (7.9)$$

Хоча у моделі HSI перетворення здійснюється за допомогою меншої кількості операцій, обчислення, необхідні для переходу з простору RGB або CMY(K) в простір HSI, не просто нівелюють (в даному конкретному випадку) цю перевагу, але й роблять такий спосіб обчислення абсолютно неефективним. Обчислювальна складність переходу у простір HSI значно перевищує складність самого перетворення. Результат перетворення, однак, не залежить від обраної для його реалізації колірної системи.

Важливо, що перетворення кожної з компонентів колірному простору за формулами (7.7) - (7.9) залежить тільки від однієї цієї компоненти. Наприклад, червона компонента s_1 на виході, відповідно (7.8), залежить тільки від червоної компоненти r_1 на вході та не залежить від зеленої (r_2) та синьої (r_3) компонент на вході. Перетворення такого типу відносяться до найбільш простих і часто використовуваних засобів обробки кольорів і можуть бути здійснені незалежно для кожної окремої колірної компоненти.

Колірним доповненням називають перехід від даних кольорів до відповідних додаткових кольорів (від червоного, зеленого, синього до жовтого, блакитного, пурпурового і навпаки). Ця операція є аналогічною перетворенню напівтонового зображення на негатив і дозволяє виявляти деталі всередині темних областей кольорового зображення (особливо коли розміри областей помітно перевищують розміри деталей). Обчислене колірне доповнення нагадує звичайний кольоровий негатив, що

отримується на кольоровій фотоплівці. Червоні кольори на початковому зображенні у колірному додатку замінюються на блакитні, біле зображення - на чорне тощо.

Вирізання певного діапазону кольорів на зображенні застосовується для виділення деяких об'єктів із їхнього оточення. Це потрібно для того, щоб або відтворити кольори, що цікавлять так, щоб вони виступали на загальному тлі, або використовувати обумовлені кольором області як маску при подальшій обробці. Найбільш просто це зробити за принципом вирізування діапазону яскравості. Однак, оскільки значення кольорового пікселя є n -вимірним вектором, функції перетворення для вирізування колірних діапазонів є більш складними, ніж аналогічні функції для вирізування діапазону яскравості. На практиці такі перетворення переважають за складністю всі колірні перетворення, що розглядалися раніше. Це пов'язано з тим, що будь-якому методі вирізування колірних діапазонів кожна колірна компонента перетвореного пікселя залежить від усіх n колірних компонент вихідного пікселя.

Один із найпростіших способів виділити певний колір зображення полягає в тому, щоб відобразити всі кольори, що лежать поза областю інтересу, в деякий нейтральний колір, що не кидається в очі. Якщо кольори, які потрібно виділити, уміщаються в деякому кубі (або гіперкубі при $n > 3$) з довжиною ребра W і з центром в точці (a_1, a_2, \dots, a_n) колірного простору, яка відповідає деякому заданому кольору-прототипу, то сукупність необхідних функцій перетворення задається виразом

$$s_i = \begin{cases} 0,5, & \text{якщо } |r_i - a_i| > W/2 \text{ для } 1 \leq j \leq n; \\ r_i, & \text{в інших випадках,} \end{cases}$$

де $i=1, 2, \dots, n$.

Ці перетворення виділяють кольори навколо заданого, замінюючи решту на колір середньої точки використовуваного колірного простору (довільно обраний нейтральний колір). Наприклад, у разі колірного простору RGB такою точкою є середина відрізка сірих кольорів, тобто колір $(0,5, 0,5, 0,5)$.

Головною перевагою каліброваних систем обробки зображень є те, що вони дозволяють здійснювати яскравість і колірну корекцію в

інтерактивному режимі незалежно, у вигляді двох послідовних операцій. Спочатку зазвичай проводиться корекція діапазону яскравості, а потім усувається колірний дисбаланс (такий як недостатня або надмірна насиченість кольорів). *Еквалізація гістограми* використовується для отримання зображення із рівномірною гістограмою значень яскравості. У випадку монохромних зображень ця процедура є дуже ефективною для якісної передачі темних, сірих та світлих ділянок зображення. Однак, оскільки кольорові зображення складаються з декількох компонентів, то не можна безпосередньо використовувати напівтонову техніку для обробки більш ніж однієї компоненти та/або гістограми. Очевидно, вирівнювання гістограми кожної компоненти кольорового зображення може призвести до появи неправильних кольорів. Більш логічний підхід полягає в однорідному розтягуванні значень інтенсивності кольорів, що не змінює самі кольори (тобто значення колірного тону). Для реалізації такого підходу ідеально пристосований колірний простір HSI.

Згладжування кольорових зображень подібно до чорно-білих напівтонових можна розглядати як операцію просторової фільтрації, при якій всі коефіцієнти фільтруючої маски мають однакові значення. По ходу переміщення маски по зображенню, що згладжується, значення кожного пікселя замінюється середнім значенням пікселів в околиці, яка визначається маскою. Основна різниця полягає в тому, що замість скалярних значень яскравості необхідно оперувати компонентами векторів кольорів. Згладжування методом усереднення по околиці може бути здійснено окремо у кожній колірній площині (по кожній колірній компоненті). Отриманий результат при цьому збігається з результатом векторного усереднення, що виконується RGB-просторі.

Аналогічним є *підвищення різкості* кольорових зображень. Традиційно для такої процедури використовується лапласіан. Вектор лапласіана визначається як вектор, компоненти якого є лапласіанами від кожної окремої скалярної компоненти вхідного вектора. Це означає, що обчислення лапласіана кольорового зображення може бути здійснено шляхом обчислення лапласіана окремо для кожної компоненти зображення.

У деяких галузях, наприклад, у дистанційному зондуванні землі

стандартною є сегментація зображення, яка полягає у розбитті зображення на окремі області. У кольорових зображеннях цю операцію виконують окремо для компонентів. Трудомісткість і результат процедури залежать від кольорового простору, що використовується. Зручно використовувати простір HSI, оскільки в ньому основна інформація про колір міститься в компоненті кольорового тону. Компонент насиченості зазвичай використовується для формування маски, що дозволяє надалі виділити області інтересу в компоненті кольорового тону. Компонента інтенсивності при сегментації кольорових зображень використовується рідше, оскільки вона не містить кольорової інформації. Проте отриманий результат сегментації у такому випадку буває не завжди якісний. Зазвичай найкращі результати досягаються під час роботи у кольоровому просторі RGB. Сегментація об'єктів на RGB-зображенні полягає у визначенні пікселів, колір яких лежить у певному діапазоні. Завдання сегментації полягає в тому, щоб класифікувати кожен піксель цього зображення відповідно до того, чи попадає його колір у заданий діапазон чи ні. Для того, щоб виконати таке порівняння, необхідно мати в кольоровому просторі певну міру подібності. Найпростішою такою мірою є евклідова відстань.

Для сегментації зображень, підвищення його різкості, стиснення важливим інструментом є *виявлення контурів*. Оконтурювання шляхом обчислення градієнта, що добре працює для напівтонових зображень, у разі окремої обробки кольорових компонентів та їх подальшого об'єднання може призвести до невірних результатів. Тому розроблені різні градієнтні методи, що застосовуються до векторних функцій.

Оскільки кількість бітів, необхідних для представлення кольору, зазвичай в три або чотири рази перевищує кількість бітів, що використовується для представлення напівтонів, то важливу роль при зберіганні та передачі кольорових зображень відіграє *стиснення даних*. Даними, що підлягають стисненню в RGB-, CMY(K)- та HSI-зображеннях, є значення компонент кожної точки кольорового зображення (наприклад, червона, зелена та синя компоненти пікселів RGB-зображення). Стиснення полягає в зменшенні або усуненні надлишкових та/або несуттєвих даних.

Насамкінець потрібно зазначити, що існують численні способи прямої

обробки кольорних векторів, які включають такі операції, як медіанна фільтрація, адаптивна фільтрація, морфологічна фільтрація, відновлення, стиснення та багато інших. Ці операції не є еквівалентними кольорній обробці, що виконується окремо для кожної компоненти кольорового зображення.

Питання для самоперевірки

1. Як можна класифікувати методи обробки зображень за призначенням?
2. Які основними стадії цифрової обробки зображень ви знаєте?
3. В чому полягає переведення зображення із аналогової в цифрову форму?
4. Які основні задачі треба вирішити при побудові квантувача зображення?
5. Якими величинами можна характеризувати колір?
6. В чому полягає фізичний сенс світлоти, кольорного тону, насиченості?
7. Сформулюйте закони змішування кольорів Грасмана.
8. В чому полягають принципи побудови колориметричних систем?
9. Які дії необхідно виконати для створення колориметричної системи?
10. Опишіть колориметричну систему RGB.
11. Опишіть колориметричну систему XYZ.
12. Опишіть колориметричну систему $L^*a^*b^*$.
13. Опишіть колориметричну системи СМУ та СМУК.
14. Які є і в чому полягають підходи, що використовуються при обробці кольорових зображень?
15. В чому полягає перетворення кольорів?
16. В чому полягають колірне доповнення, вирізання діапазону кольорів?
17. В чому полягають еквалізація гістограми, згладжування кольорових зображень?

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ДСТУ 2681-94. Метрологія. Терміни та визначення
2. Величко О. М. Оцінювання результатів вимірювань: основи і нормативне забезпечення / О. М. Величко, Л. В. Коломієць, Т. Б. Гордієнко. – Одеса: ВМВ, 2010. – 380 с.
3. Сато Ю. Обробка сигналів. Перше знайомство / Ю. Сато. – Додека ХХІ, 2008. – 176 с.
4. Братченко Г. Д., Перелигін Б. В., Банзак О. В., Казакова Н. Ф., Григор'єв Д. В. Методи та засоби обробки сигналів. Навчальний посібник. – Одеса: Типографія-видавництво «Плутон», 2014. – 452 с.
5. Richard G. Lyons Understanding Digital Signal Processing, 3rd Edition. Pearson, 2010.
6. Бабак В.П., Хандлецький В.С., Шрюфер Е. Обробка сигналів: Підручник. – К.: Либідь, 1996. – 392 с.
7. Алмазов В. Б. Основы теории радиолокации / В. Б. Алмазов. – Харьков: ВИРТА им. Говорова Л. А., 1992. – 204 с.
8. Mallat S. A Wavelet Tour of Signal Processing. The Sparse Way. Third Edition. Academic Press. 2009
9. Blatter C. Wavelets: A Primer. CRC Press, 2018
10. Основи та методи цифрової обробки сигналів: від теорії до практики: навч. посібник / Ушенко Ю.О., М.С. Гавриляк, М.В. Талах, В.В. Дворжак. – Чернівці : Чернівецький нац. ун-т ім. Ю. Федьковича. – 2021. – 308 с.
11. Сергієнко А. Б. Цифрова обробка сигналів / А. Б. Сергієнко. – СПб.: Пітер, 2002. – 608 с. (російською)
12. Rafael C. Gonzalez, Richard E. Woods. Digital Image Processing, 4th edition. 2018