Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Інститут електроенергетики
Факультет інформаційних технологій
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеню бакалавра

студента *Ковальська Олена Олександрівна*

академічної групи *172-16зск-1*

спеціальності *172 Телекомунікації та радіотехніка*

спеціалізації[1]

за освітньо-професійною програмою *Телекомунікації та радіотехніка*

на тему *Удосконалення способу мовного кодування в телекомунікаційній системі*

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
Кваліфікаційної роботи	к.ф.-м..н., доц. Магро В.І.			
розділів:				
спеціальний	к.ф.-м..н., доц. Магро В.І.			
економічний	к.е.н., доц. Романюк Н.М.			
Рецензент				
Нормоконтролер	к.ф.-м.н., проф. Гусєв О.Ю.			

Дніпро
2019

ЗАТВЕРДЖЕНО:
завідувач кафедри
безпеки інформації та телекомунікацій
_____ д.т.н., проф. Корнієнко В.І.

«_____»_____20___року

ЗАВДАННЯ
на кваліфікаційну роботу
ступеня бакалавра

студентці ***Ковальській Олені Олександрівні*** **академічної групи** ***172-16зск-1***
(прізвище ім'я по-батькові) (шифр)

спеціальності *172 Телекомунікації та радіотехніка*
(код і назва спеціальності)

на тему ***Удосконалення способу мовного кодування в телекомунікаційній системі***

затверджену наказом ректора НТУ «Дніпровська політехніка» від _____ № ____

Розділ	Зміст	Термін виконання
Розділ 1	Розглянути методів мовного кодування та порядок перетворення сигналу в цифрову форму	
Розділ 2	Удосконалити спосіб підвищення якості мовного сигналу без збільшення швидкості передачі	
Розділ 3	Провести розрахунок витрат на розробку нового способу кодування	

Завдання видано _____ **Магро В.І.**
(підпис керівника) (прізвище, ініціали)

Дата видачі: _____

Дата подання до екзаменаційної комісії: _____

Прийнято до виконання _____ **Ковальська О.О.**
(підпис студента) (прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 87 с., 38 рис., 3 табл., 1 додаток, 15 джерел.

Об’єкт дослідження: метод цифрового перетворення мовного сигналу.

Мета роботи: підвищення якості мовного сигналу без збільшення швидкості передачі за рахунок реконфігурації мультисмугової кодової книги.

Метод дослідження: технічний з використанням системного аналізу.

В першому розділі кваліфікаційної роботи розглянуто методи мовного кодування, механізм мовоутворення. Детально розглянуті схеми кодування мовних сигналів.

В спеціальній частині розглядається ДПФ з нерівномірним частотним дозволом, що дозволяє апроксимувати психоакустичні шкали для побудови перцептуально мотивованих систем стиснення і передачі мовної інформації, а також систем підвищення якості мови. Виконана реалізація фільтрів в середовищі MATLAB.

В економічному розділі проведено розрахунок витрат на розробку нового способу кодування.

Практичне значення отриманих результатів полягає у вдосконаленні методу кодування лінійного передбачення з мультикодовим управлінням. Запропонована модель кодера, яка дозволяє знизити об'єм переданої інформації, зберігаючи високу якість мовного сигналу. Після оптимізації об’єму кодових книг різних рівнів з урахуванням апріорних знань про перцептуальну значимість субсмуг якість закодованого сигналу підвищилася на 0,24 бала за п'ятибальною шкалою MOS.

ПЕРЦЕПТУАЛЬНО МОТИВОВАНА СИСТЕМА СТИСНЕННЯ. БАНК ЦИФРОВИХ ФІЛЬТРІВ. WARPED DISCRETE-FOURIER TRANSFORM. СИСТЕМА МУЛЬТИМЕДІА.

РЕФЕРАТ

Пояснительная записка: 87 с., 38 рис., 3 табл., 1 приложение, 15 источников.

Объект исследования: метод цифрового преобразования речевого сигнала.

Цель работы: повышение качества речевого сигнала без увеличения скорости передачи за счет реконфигурации мультиполосной кодовой книги.

Метод исследования: технический с использованием системного анализа.

В первом разделе квалификационной работы рассмотрены методы языкового кодирования, механизм речепорождения. Подробно рассмотрены схемы кодирования речевых сигналов.

В специальной части рассматривается ДПФ с неравномерным частотным разрешением, позволяющий аппроксимировать психоакустические шкалы для построения перцептуально мотивированных систем сжатия и передачи речевой информации, а также систем повышения качества речи. Выполнена реализация фильтров в среде MATLAB.

В экономическом разделе проведен расчет затрат на разработку нового способа кодирования.

Практическое значение полученных результатов заключается в совершенствовании метода кодирования линейного предсказания с мультикодовим управлением. Предложенная модель кодера, которая позволяет снизить объем передаваемой информации, сохраняя высокое качество речевого сигнала. После оптимизации объема кодовых книг разных уровней с учетом априорных знаний о перцептуальной значимости субполос качество закодированного сигнала повысилась на 0,24 балла по пятибалльной шкале MOS.

ПЕРЦЕПТУАЛЬНО МОТИВИРОВАННАЯ СИСТЕМА СЖАТИЯ. БАНК ЦИФРОВЫХ ФИЛЬТРОВ. WARPED DISCRETE-FOURIER TRANSFORM. СИСТЕМА МУЛЬТИМЕДИА.

ABSTRACT

Explanatory note: 87 p., 38 figures, 3 tables, 1 supplement, 15 sources.

Object of research:- the method of digital conversion of speech signal.

The purpose of the work is to improve the quality of the speech signal without increasing the transmission speed by reconfiguring the multistage code book.

Research method: technical using system analysis.

In the first section of the qualification paper, the methods of speech coding, the mechanism of language formation are considered. Detailed coding schemes for speech signals are considered in detail.

In the special section is considered DFT with uneven frequency resolution, which allows to approximate the psychoacoustic scales for the construction of perceptually motivated compression and transmission of speech information, as well as systems for improving the quality of speech. Realization of filters in MATLAB environment is realized.

In the economic section, the cost of developing a new method of coding are calculated.

The practical value of the results obtained is to improve the linear prediction encoding method with multi-code control. The proposed model of the encoder, which allows to reduce the amount of transmitted information, while maintaining the high quality of the speech signal. After optimizing the volume of code books of different levels, taking into account a priori knowledge about the perceptual significance of sub bands, the quality of the encoded signal increased by 0.24 points on a five-point MOS scale.

PERCEPTUALLY MOVED SYSTEM OF STABILITY. BANK OF DIGITAL FILTERS. WARPED DISCRETE-FOURIER TRANSFORM. MULTIMEDIA SYSTEM.

СПИСОК УМОВНИХ СКОРОЧЕНЬ

АДІКМ – адаптивна диференціальна імпульсно-кодова модуляція;

БФ – банк фільтрів;

ВІС – велика інтегральна схема;

ДІКМ – диференціальна імпульсно-кодова модуляція;

ДПФ – дискретне перетворення Фур'є;

ІКМ – імпульсно-кодова модуляція;

КІХ – кінцева імпульсна характеристика;

ЛПК – лінійні передбачувані кодери;

CELP – codebook excited linear prediction;

CELP – кодер зі збудженням на основі кодової книги

MOS – mean opinion score;

MOS – усереднена оцінка розбірливості мови;

MPE – multi-pulse excited;

RPE – regular pulse excited;

WDFT – дискретне перетворення Фур'є з нерівномірним розкладанням коефіцієнтів Фур'є по частотній вісі;

WDFT – warped discrete Fourier transform.

ЗМІСТ

ВСТУП…….....9
1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ……10
1.1Огляд методів мовного кодування. Механізм мовоутворення……..10
1.2 Перетворення сигналу в цифрову форму……13
1.2.1 Квантувачі з рівномірним кроком……13
1.2.2 Квантувачі з нерівномірним кроком……..16
1.3 Класифікація схем кодування мовних сигналів……..17
1.3.1 Кодування форми сигналу…….18
1.3.2 Кодування джерела…….21
1.3.3 Канальні вокодери……23
1.3.4 Гомоморфні вокодери…….24
1.3.5 Вокодери з лінійним передбаченням…….25
1.4 Гібридні методи кодування мови…….26
1.4.1 Кодери на основі «Аналіз через синтез»……..27
1.4.2 Кодери з постійними імпульсними збудженнями……28
1.4.3 Кодери зі збудженням на основі кодових книг……..29
1.5 Параметри порівняння методів кодування мови…….33
1.6 Висновок до першого розділу……37
1.6 Постановка задачі до другого розділу……38
2 СПЕЦІАЛЬНА ЧАСТИНА……40
2.1 Банки цифрових фільтрів…….....40
2.1.1 ДПФ-модульований рівносмуговий банк фільтрів на основі поліфазної структури фільтрів дециматорів і інтерполяторів ……47
2.1.2 ДПФ – модульований нерівносмуговий банк фільтрів……50
2.1.3 Приклад банку ДПФ – модульованих поліфазних фільтрів……..54
2.2 Косинусно – модульовані банки фільтрів ……....55
2.2.1 Проектування косинусно – модульованого банку фільтрів……55
2.2.2 Косинусно – модульовані банки фільтрів з квазиповним відновленням.57
2.3 ДПФ з нерівномірною частотною роздільністю (WDFT)…….....58
2.3.1 Основні принципи WDFT…….....58

2.3.2 Зворотнє перетворення WDFT…….....60
2.3.3 Оцінка помилки реконструкції WDFT…….....63
2.3.4 Формування надлишкового базису WDFT……65
2.3.5 Апроксимація психоакустичних шкал…….....66
2.4 Побудова моделі кодування мови із застосуванням банків цифрових фільтрів і WDFT……69
2.4.1 Опис узагальненої моделі кодування мови…….....69
2.4.2 Побудова моделі перцептуального кодера мови. Структура кодера……71
2.4.3 Розбиття частотної осі на субсмуги……72
2.4.4 Реконфігурація структури кодової книги і математичне моделювання в пакеті MATLAB ……74
2.5 Висновок до другого розділу……..79
3 ЕКОНОМІЧЧНИЙ РОЗДІЛ……80
3.1 Техніко – економічне обґрунтування створення нового алгоритму кодування мови……80
3.2 Визначення трудомісткості розробки алгоритму……..80
3.3 Витрати на створення алгоритму кодування……82
3.4 Визначення економічної переваги розробленого алгоритму…….84
ВИСНОВКИ…….85
ПЕРЕЛІК ПОСИЛАНЬ……86
ДОДАТОК А
ДОДАТОК Б
ДОДАТОК В
ДОДАТОК Г

ВСТУП

Сучасний етап розвитку суспільства характеризується збільшенням потоків мовних повідомлень, що передаються по телекомунікаційним системам та мережам, які мають обмежену пропускну спроможність. Великі труднощі виникають також при вирішенні задачі зберігання мовних даних, зростаючий потік яких потребує все більших обсягів пам'яті запам'ятовуючих пристроїв.

У різних країнах світу, для вирішення цієї проблеми інтенсивно розвивається напрям, що полягає в стисненні обсягів мовних даних. Цей напрямок особливо привабливо з економічної точки зору, так як не вимагає створення нових дорогих виробничих потужностей, але призначено для ефективного використання наявних ресурсів систем і мереж.

Один з перспективних напрямів є побудова широкосмугових CELP кодерів мови. Вони задовольняють двом основним вимогам: відрізняються дешевизною у виконанні та мають достатню надійність від злому, що забезпечується використанням шумоподібних сигналів, які за своїми властивостями, практично ідентичні білому гауссовському шуму.

Під час мовоутворення слід враховувати психоакустичну модель сприйняття звуків людиною. згідно цієї моделі існують домінантні частотні смуги в мовному сигналі які формують перцептуальне (чуттєве) комфортне сприйняття звуку людиною. Сигналі з інших частотних смуг не справляють суттєвого впливу на якість сприйнятого сигналу. Отже, сигнали що припадають на інші частотні смуги можуть бути видалені, тобто об’єм мовної інформації зменшиться.

Всі зазначені вище чинники зумовлюють актуальність теми кваліфікаційної роботи.

Метою кваліфікаційної роботи є підвищення якості мовного сигналу без збільшення швидкості передачі за рахунок реконфігурації мультисмугової кодової книги.

1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

1.1 Огляд методів мовного кодування. Механізм мовоутворення

Значний об’єм переданої інформації в системах зв'язку сьогодні припадає на мову – це дротова телефонія, системи стільникового і супутникового зв'язку і т.д. Тому ефективному кодуванню, або стисненню мови, в системах зв'язку приділяється виняткова увага.

Розглянемо основні властивості мовного сигналу як об'єкта кодування і передачі по каналах зв'язку і спробуємо пояснити, на яких властивостях сигналу грунтується можливість його стиснення.

Мова являє собою коливання складної форми, яка залежить від вимовлених слів, тембру голосу, інтонації, статі і віку мовця. Спектр мови досить широкий (приблизно від 50 до 10000 Гц), але для передачі мови в телефонії відмовилися від складових, що лежать за межами смуги частот 0,3 - 3,4 кГц, що дещо погіршило сприйняття ряду звуків (наприклад, шиплячих, істотна частина енергії яких зосереджена у верхній частині мовного спектру), але мало торкнулося розбірливості.

Обмеження частоти знизу (до 300 Гц) також трохи погіршує сприйняття через втрати низькочастотних гармонік основного тону.

Другою особливістю мовних сигналів є нерівномірність розподілу ймовірностей (щільності ймовірності) миттєвих значень сигналу. Малі рівні сигналу значно більш вірогідні, ніж великі.

Особливо це помітно на фрагментах великої тривалості з невисокою активністю мови. Цей фактор, також забезпечує можливість економного кодування - більш ймовірні значення можуть кодуватися короткими кодами, менш імовірні - довгими.

Ще одна особливість мовних сигналів - їх істотна нестаціонарність у часі: властивості і параметри сигналу на різних ділянках значно різняться. При цьому розмір інтервалу стаціонарності складає порядку декількох

десятків мілісекунд. Ця властивість сигналу значно ускладнює його економне кодування і змушує робити системи стиснення адаптивними, тобто підлаштовується під значення параметрів сигналу на кожній з ділянок.

Нарешті, виключно важливим для організації стиснення мовних сигналів є розуміння фізики механізму мовоутворення. Мова формується при проходженні виштовхування легким потоком повітря через голосові зв'язки і голосовий тракт.

Голосовий тракт починається від голосових зв'язок і закінчується губами і в середньому має довжину близько 15-17 сантиметрів. Голосовий тракт в силу своїх резонансних властивостей вносить у сформований сигнал набір характерних для кожної людини частотних складових, званих формантами, що додають голосу тембрового забарвлення.

Частоти і полоси формант можуть управлятися зміною форми голосового тракту, наприклад, зміною положення мови.

Важливою частиною багатьох голосових кодерів є моделювання голосового тракту як короткочасного фільтра із змінними параметрами.Оскільки форма голосового тракту може змінюватися порівняно повільно (важко припустити, що можна змінювати положення мови частіше, ніж 20 - 30 разів на секунду), то параметри такого фільтра повинні оновлюватися чи змінюватися) також порівняно рідко (зазвичай - через кожні 20 мілісекунд або навіть рідше).

Таким чином, голосовий тракт збуджується потоком повітря, які направляються у нього через голосові зв'язки.

Залежно від способу порушення виникаючі при цьому звуки можна умовно розділити на три класи:

1) Голосні звуки, що виникають, коли голосові зв'язки вібрують, відкриваються і закриваються, перериваючи тим самим потік повітря від легких до голосового тракту. Збудження голосового тракту при цьому проводиться квазіперіодичними імпульсами. Швидкість (частота) відкривання і закривання зв'язок визначають висоту виникнення звуку (тону).

Вона може управлятися зміною форми і напруги голосових зв'язок, а також зміною тиску повітряного потоку, який підходить. Голосні звуки мають високий ступінь періодичності основного тону з періодом 2 - 20 мс.

2) Відповідно звуки, що виникають при порушенні голосового тракту шумоподібним турбулентним потоком, який формується проходять з високою швидкістю через відкриті голосові зв'язки потоком повітря. У таких звуках практично відсутня довгострокова періодичність, обумовлена вібрацією голосових зв'язок.

3) Звуки вибухового характеру, що виникають, тоді коли закритий голосової тракт з надлишковим тиском повітря раптово відкривається. Деякі звуки в чистому вигляді не підходять ні під один з описаних вище класів, але можуть розглядатися як їх суміш. Таким чином, процес мовоутворення можна розглядати як фільтрацію мовоутворюючим трактом зі змінюваними в часі параметрами сигналів збудження,а також зі змінюваними характеристиками. Модель процесу мовоутворення показана на рисунку 1.1.

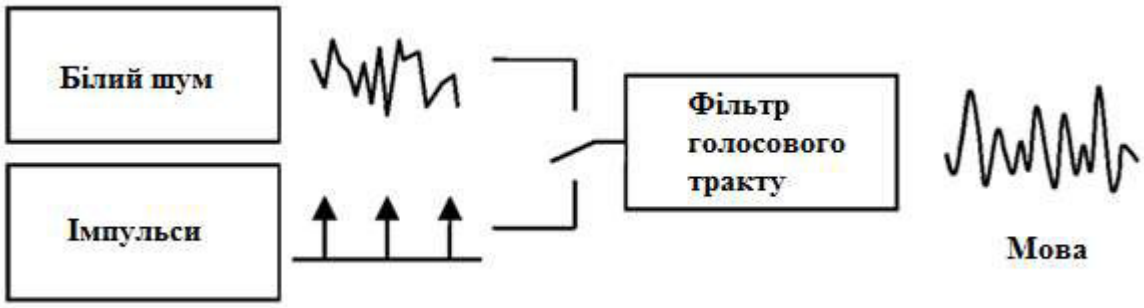


Рисунок 1.1 – Модель процесу мовоутворення

При цьому, незважаючи на виняткову різноманітність генерації мовних сигналів, форма і параметри голосового тракту, а також способи і параметри збудження досить одноманітні і змінюються порівняно повільно.

Більшість кодерів мови використовують цю передбаченність, а також низьку швидкість зміни параметрів моделі системи мовоутворення для зменшення швидкості коду.

Однією з обов'язкових ланок системи обробки мовних даних є квантувач, який дозволяє перетворити сигнал з аналогової форми в цифрову.

1.2 Перетворення мовного сигналу в цифрову форму

1.2.1 Квантувачі з рівномірним кроком

Безперервний у часі та змінюваний за рівнем мовний сигнал повинен бути перетворений в дискретний сигнал. На першому етапі в дискретизаторі (рис. 1.2) аналоговий сигнал з обмеженим спектром замінюють безліччю його наступних відліків , з постійним інтервалом, через T_s секунд. Однак за амплітудою ці миттєві відліки не квантовані.

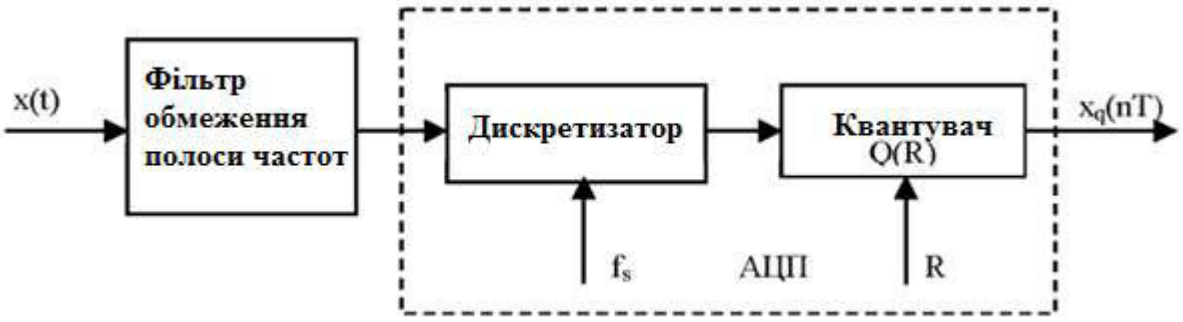


Рисунок 1.2 – Перетворення мовного сигналу в дискретний

Дискретизація заснована на теоремі Котельникова, згідно з якою сигнал з обмеженим спектром частот можна повністю представити сукупністю відліків з тактовою частотою $f_s = 1/T_s$, яка принаймні вдвічі перевищує найвищу частоту спектру сигналу.

На вході дискретизатора необхідний фільтр нижніх частот, що забезпечує придушення спектральних компонентів сигналу з частотами, більшими $f_s/2$.

Якщо такі компоненти потраплять на дискретизатор, то виникне характерний вигляд спотворень, пов'язаний з перекриттям спектрів власне сигналу і сигналу, перенесеного на f_s.

Необхідність надійного захисту від перекриття спектрів обумовлює вибір значення f_s, що перевищує подвійне значення верхньої частоти сигналу.

Наприклад, при дискретизації стандартного телефонного сигналу з полосою 300 - 3400 Гц, частоту дискретизації вибирають 8 кГц, а не 6.8 кГц.

Перевищення значенням f_s подвоєного максимальною частотою спектру сигналу називають надлишковою дискретизацією.

Другий етап – квантування за рівнем, що виконується квантувачем. При цьому безперервний інтервал значень переданого сигналу замінюють кінцевим, безліччю дозволених для передачі значень.

При рівномірному квантуванні весь діапазон значень переданого сигналу $x_{max} – x_{min}$ розбивають на $N=2^R$ рівнів, віддалених один від одного на крок квантування D, так, що:

$$x_{max} - x_{min} = N*\Delta \qquad (1.1)$$

де x_{max} – максимальне значення вхідної величини;

x_{min} – мінімальне значення вхідної величини;

N – число рівнів квантування;

A – крок квантування.

Значення R прийнято називати розрядністю квантувача. Наприклад, дуже широко поширене восьмирозрядне або восьмибітове (R = 8) квантування, при якому можливі значення сигналу співвідносять з одним з 256 рівнів. Процес дискретизації сигналу за часом і квантування його за рівнем пояснює рис. 1.3.

Третій етап, також виконується квантувачем - кодування квантованих значень відліків набором двійкових символів, званих кодовим словом.

Найбільш просто кодування можна здійснити, записавши номер рівня квантування у вигляді двійкового числа. Наприклад, рівень 57 з 256 дозволених рівнів буде представлений восьмирозрядним двійковим числом 00111001.

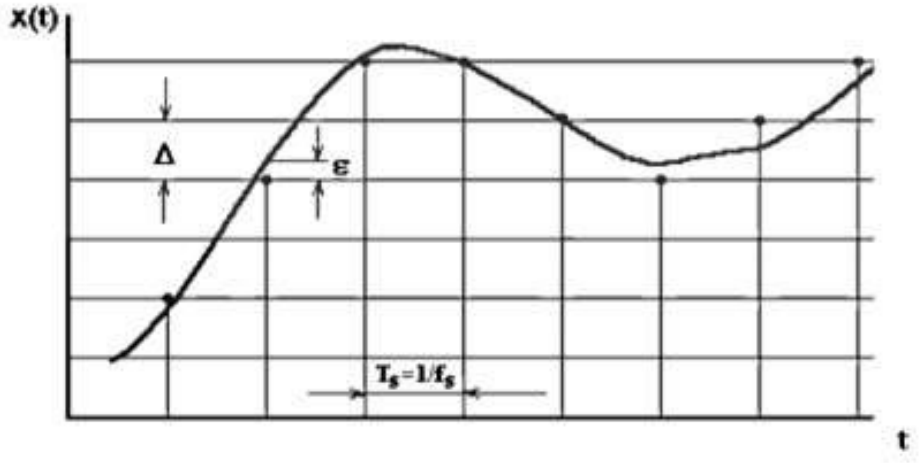


Рисунок 1.3 – Дискретизація сигналу за часом і квантування

Таке уявлення є найпростішим, однак, не завжди є найзручнішим. При квантуванні виникає помилка квантування e(n), рівна різниці між квантованим ($x_q(n)$) і істинним (x(n)) значенням сигналу :

$$s(n) = x(n) - x_q(n) \tag{1.2}$$

Усередині динамічного діапазону помилка обмежена:

$$\varepsilon(n) \leq \frac{\Delta}{2} \tag{1.3}$$

де Δ – крок квантування

За межами динамічного діапазону, коли квантувач насичений, помилка зростає пропорційно x(n). Рис. 1.3 ілюструє утворення помилок квантування при квантуванні з рівномірним кроком.

Спотворення, що вносяться квантувачем, оцінюють значенням середньоквадратичної помилки. Середньоквадратичну помилку визначають як:

$$D_{cko} = \sqrt{\frac{1}{N}} * \sum_{i=1}^{N} (x(i)) - x_q(i))^2 \tag{1.4}$$

де D_{cko} – середньоквадратична помилка квантування.

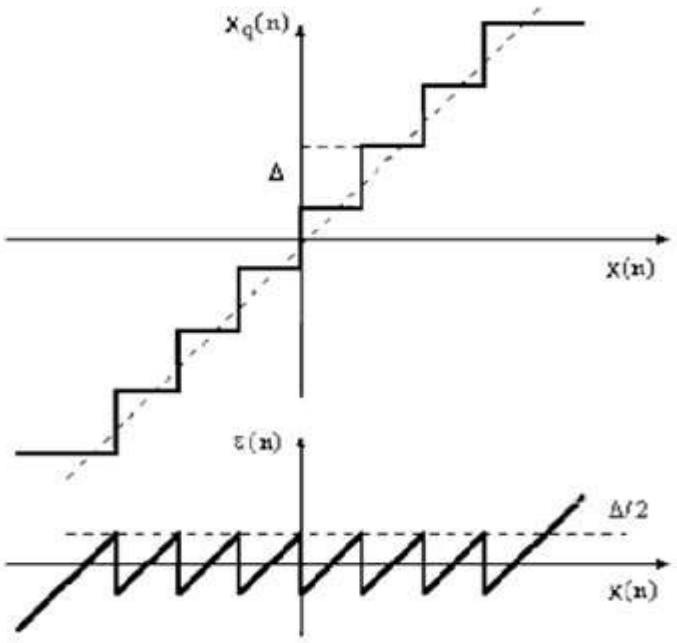


Рисунок 1.4 – Помилки при квантуванні з рівномірним кроком

Як випливає з виразу (1.4) для зниження помилки квантування необхідно знижувати крок квантування і відповідно збільшувати число рівнів квантування.

1.2.2 Квантувачі з нерівномірним кроком

При рівномірному квантуванні помилка квантування найбільш помітна при малих рівнях сигналу, так як зменшення потужності переданого сигналу призводить до зниження відношення сигнал-шум. Однак збільшення потужності сигналу також погіршує відношення сигнал-шум через наступаюче при цьому перевантаження.

У сучасних системах використовують нерівномірне квантування, при якому малі рівні сигналу кванти з меншою помилкою, ніж великі. Змінюючи крок квантування пропорційно рівню вхідного сигналу, відношення сигнал-шум зберігають постійним при зміні рівня сигналу.

При компресії роблять нелінійне перетворення сигналу для того, щоб вирівняти густину ймовірності розподілу миттєвих значень, перетворюючи

тим самим реальний сигнал в сигнал з рівномірним розподілом миттєвих значень. Такий сигнал є оптимальним для квантувача з рівномірним кроком.

Умовно нерівномірне квантування можна представити як послідовне з'єднання пристрою компресії вхідного сигналу і рівномірного квантувача (рис. 1.5).

При прийомі нелінійні спотворення сигналу, що вносяться компресором, усуваються експандером, нелінійним пристроєм з амплітудною характеристикою, зворотній характеристиці компресора.

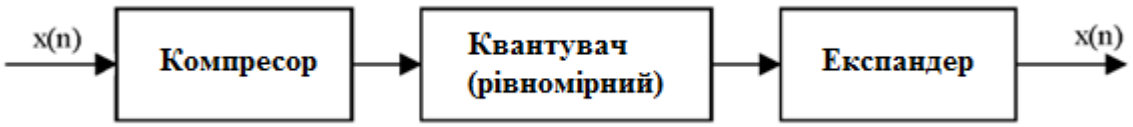


Рисунок 1.5 – Система з нерівномірним квантуванням

Процес компресії і експандування називають компандуванням. При передачі мовних даних найчастіше використовують два типи компандування: по μ-закону и по А-закону.

1.3 Класифікація схем кодування мовних сигналів

Основне завдання і функція схем кодування - проаналізувати вхідний сигнал, видалити надмірність і відповідним чином закодувати інформативні частини сигналу. Для зниження швидкості цифрового потоку доводиться розробляти все більш складні методи усунення надмірності.

Мовне кодування передбачає перетворення мовного аналогового аудіосигналу в цифровий сигнал з подальшим усуненням надлишкової інформації для передачі по каналах зв'язку, запису або подальшому перетворенню в ефективний формат.

Мовні кодери можна розділити на дві великі групи: кодери форми або звичайні кодери і вокодери, призначені виключно для передачі мови (рис. 1.6).

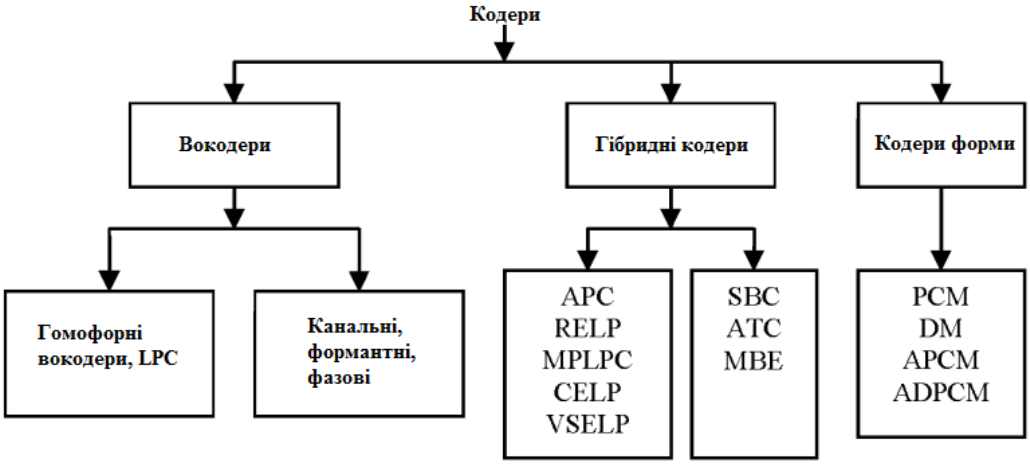


Рисунок 1.6 – Классифікація схем кодування

1.3.1 Кодування форми сигналу

Найпростіші кодери мови, взагалі не використовують інформацію про те, як був сформований закодований сигнал, а просто намагаються максимально наблизити відновлений сигнал за формою до оригіналу,такими є кодери форми сигналу. Теоретично вони інваріантні до характеру сигналу, що подається на їх вхід, і можуть використовуватися для кодування будь-яких, в тому числі і немовних, сигналів**.**

Ці кодери – найпростіші за принципом дії, але великих ступенів стиснення (низьких швидкостей коду) забезпечити не можуть.

Найпростішим способом кодування форми сигналу є так звана імпульсно-кодова модуляція (ІКМ), при використанні якої виробляються просто дискретизація і рівномірне квантування вхідного сигналу, а також перетворення отриманого результату в рівномірний двійковий код.

Для мовних сигналів зі стандартною полосою частот для передачі мови 0,3 - 3,4 кГц зазвичай використовують частоту дискретизації $F_{дискр}>2F_{max}$ і рівну 8 кГц.

Експериментально показано, що при рівномірному квантуванні для отримання практично ідеальної якості мови потрібно квантувати сигнал не менше ніж на (плюс/мінус) 2000 рівнів, іншими словами, для подання кожного відліку знадобиться 12 біт, а результуюча швидкість коду становитиме:

$$R = 8000 \text{ відліків/с} * 12 \text{ біт/відлік} = 96000 \text{ біт/с} = 96 \text{ кбіт/с} \quad (1.5)$$

Використовуючи нерівномірне квантування (більш точне для малих рівнів сигналу і більше грубе для великих його рівнів, таким чином, щоб відносна помилка квантування була постійною для всіх рівнів сигналу), можна досягти тієї ж самої суб'єктивної якості відновлення мовного сигналу, але при набагато меншому числі рівнів квантування - порядку (плюс/мінус) 128.

У цьому випадку для двійкового подання відліків сигналу знадобиться вже 8 біт і результуюча швидкість коду складе 64 кбіт / с.

Наступним прийомом, що дозволяє зменшити результуючу швидкість коду, може бути спроба передбачити значення поточного відліку сигналу по деяким попереднім його значенням, і далі, кодування вже не самого відліку, а помилки його передбачення - різниці між істинним значенням поточного відліку і його передбаченим значенням.

Якщо точність передбачення досить висока, то помилка передбачення чергового відліку буде значно менше величини самого відліку і для її кодування знадобиться набагато менше число біт.

Таким чином, чим більш передбачуваним буде поведінка закодованого сигналу, тим ефективнішим буде його стиснення.

Описана ідея лежить в основі так званої диференціальної імпульсно-кодової модуляції (ДІКМ) - способу кодування, при якому кодуються не власні значення сигналу, а їх відмінності від передбачених значень.

Найпростішим способом передбачення є використання попереднього відліку сигналу в якості передбачення його поточного значення. Це так зване передбачення нульового порядку, найпростіше, але і найменш точне. Більш

точним, очевидно, буде передбачення поточного відліку на основі лінійної комбінації двох попередніх і т.д.

На жаль, точність передбачення не завжди зростає з ростом порядку передбачення, оскільки властивості сигналу між відліками починають вже змінюватися, тому зазвичай обмежуються передбаченням не вище 2 - 3-го порядку. На рис. 1.7 наведена схема ДІКМ кодера.

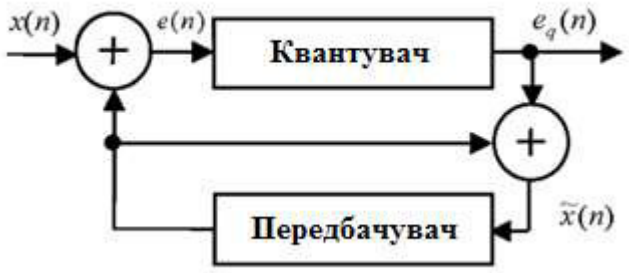


Рисунок 1.7 – Структурна схема ДІКМ кодера

У цьому випадку при кодуванні мовних сигналів з урахуванням ступеня їх короткочасної (на кілька чергових відліків) передбачуваності результуюча швидкість коду для ДІКМ зазвичай складає 5-6 біт на відлік або 40 - 48 кбіт / с.

Ефективність ДІКМ може бути підвищена, якщо передбачення і квантування сигналу буде виконуватися не на основі деяких усереднених його характеристик, а з урахуванням їх поточного значення та зміни в часі, тобто адаптивно.

Так, якщо швидкість зміни сигналу стала більшою, можна збільшити крок квантування, і, навпаки, якщо сигнал став змінюватися набагато повільніше, величину кроку квантування можна зменшити. При цьому помилка передбачення зменшиться і, отже, буде кодуватися меншим числом біт на відлік. Такий спосіб кодування називається адаптивною ДІКМ, чи АДІКМ (ADPCM).

Сьогодні такий спосіб кодування стандартизований і широко використовується при кодуванні мови в міжміських цифрових системах

зв'язку, в системі мікростільникового зв'язку DECT, в цифрових безшнурових телефонах і т.д.

Використання АДІКМ зі швидкістю коду 4 біта / відлік або 32 кбіт / с забезпечує таку ж суб'єктивну якість мови, що і 64 кбіт / с - ІКМ, але при вдвічі меншій швидкості коду.

На сьогодні стандартизовані також АДІКМ-кодери для швидкостей 40, 24 і 16 кбіт / с (в останньому випадку з дещо гіршим, ніж для 32 кбіт / с - АДІКМ, якістю сигналу).

Таким чином, видно, що стиснення мовних сигналів на основі кодування їх форми забезпечує в кращому випадку дво - триразове зменшення швидкості коду. Подальше зниження швидкості веде до різкого погіршення якості кодованого сигналу.

Наступним великим класом кодерів мовних сигналів є кодери джерела.

1.3.2 Кодування джерела

На відміну від кодерів форми сигналу, які взагалі не використовують інформацію про те, як був сформований закодований сигнал, кодери джерела грунтуються саме на моделі джерела і із закодованого сигналу витягують інформацію про параметри цієї моделі.

При цьому результатом кодування не є коди сигналів, а коди параметрів джерела цих сигналів.

Кодери джерела для кодування мови називаються вокодери і працюють таким чином.

Голосоутворюючий тракт представляється як лінійний фільтр зі змінними в часі параметрами, який збуджується або джерелом білого шуму (при формуванні приголосних звуків), або послідовностями імпульсів з періодом основного тону (при формуванні голосних звуків), рис. 1.8.

Лінійна модель системи мовоутворення і її параметри можуть бути знайдені різними способами. І від того, яким способом вони визначаються, залежить тип вокодера.

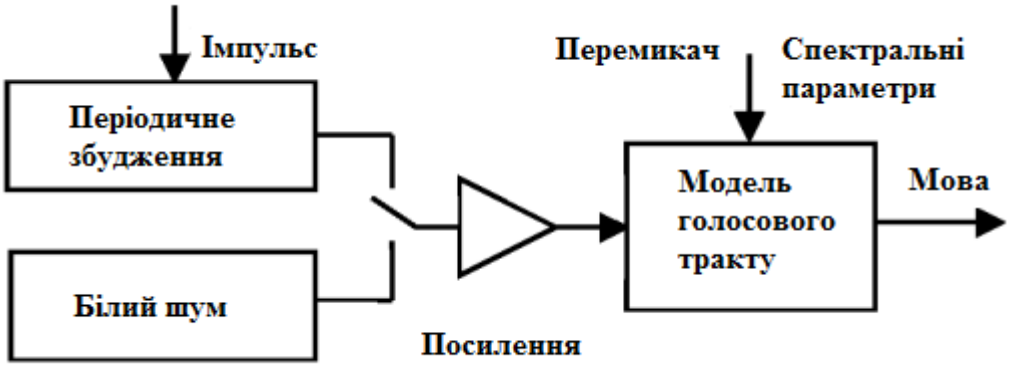


Рисунок 1.8 – Голосоутворюючий тракт при вокодерному кодуванні

Інформація, яку отримує вокодер в результаті аналізу мовного сигналу і передає декодеру, це параметри мовоутворюючого фільтра, покажчик голосних/приголосих звуків, потужність сигналу збудження і період основного тону для голосних звуків.

Ці параметри повинні оновлюватися кожні 10-20 мс, щоб відстежувати нестаціонарний характер мовного сигналу.

Вокодер, на відміну від кодера форми сигналу, намагається сформувати сигнал, що звучить як оригінальна мова, і не звертає уваги на відмінність форми цього сигналу від початкового. При цьому результуюча швидкість коду на його виході зазвичай складає не більше 4,8 кбіт/с, тобто в п'ятнадцять разів менше, ніж при ІКМ.

На жаль, якість мови,яку забезпечують вокодери, дуже далека від ідеального, її звучання хоча і досить розбірливо, але абсолютно ненатурально.

При цьому навіть істотне збільшення швидкості коду практично не покращує якості мови, оскільки для кодування була обрана занадто проста модель системи мовоутворення.

Особливо грубим є припущення про те, що мова складається лише з голосних і приголосних звуків, і не допускає будь-яких проміжних станів.

Вокодери призначені виключно для кодування мовного сигналу. При їх побудові максимально враховують особливості утвореня мови і її

сприйняття людиною. Форма відновленого сигналу при цьому може радикально відрізнятися від форми вихідного сигналу.

Основне застосування вокодери знайшли у військовій області, де головне - це не натуральність мови, а велика ступінь її стиснення і дуже низька швидкість коду, що дозволяє ефективно захищати від перехоплення і засекречувати передану мову.

Коротко розглянемо основні з відомих типів вокодерів.

1.3.3 Канальні вокодери

Це найбільш архаїчний тип вокодера, запропонований ще в 1939 році. Цей вокодер використовує слабку чутливість слуху людини до незначних фазових (часових) зрушень сигналу, оскільки органи слуху людини не реагують на фазові співвідношення.

Для сегментів мовлення довжиною приблизно в 20 - 30 мс за допомогою набору вузькополосних фільтрів визначається амплітудний спектр. Чим більше фільтрів, тим краще оцінюється спектр, але тим більше потрібно біт для його кодування і тим більша результуюча швидкість коду.

Сигнали з виходів фільтрів детектуються, пропускаються через ФНЧ, дискретизуються і піддаються бінарному кодуванню (рисунок 1.9).

Рисунок 1.9 – Канальний вокодер

Таким чином визначаються повільно змінюючі параметри мовоутворювального тракту, а також за допомогою детекторів основного тону і голосних звуків, період основного тону збудження і ознака - голосний / приголосний звук.

Канальний вокодер може бути реалізований як у цифровій, так і в аналоговій формі і забезпечує досить розбірливу мову при швидкості коду на його виході близько 4,8 кбіт / с.

Враховуючи простоту моделі, важко очікувати від вокодерного кодування хорошої якості відновленої мови. Дійсно, канальні вокодери використовуються в основному тільки там, де головним чином необхідні розбірливість і високий ступінь стиснення: у військовій зв'язку, авіації, космічному зв'язку.

Розглянемо більш нові і більш складні методи вокодерного кодування мовних сигналів.

1.3.4 Гомоморфні вокодери

Гомоморфна обробка сигналів являє собою один з нелінійних методів обробки, який може ефективно застосовуватися в складних сигналах, наприклад в мовних і дозволяє на своїй основі будувати вокодер.

Гомоморфна обробка дозволяє розділити мовний сигнал на генераторну і фільтрову частини.

З урахуванням використовуваної в вокодерах моделі системи голосоутворення мовний сигнал можна представити як тимчасову згортку імпульсної перехідної характеристики голосового тракту з сигналом збудження. У частотній області це відповідає добутку частотної характеристики голосового тракту і спектру сигналу збудження.

Якщо взяти логарифм від цього, то отримаємо суму логарифмів спектру сигналу збудження і частотної характеристики голосового тракту.

Оскільки людське вухо практично не відчуває фази сигналу, можна оперувати амплітудними спектрами:

$$\lg(|S(e^{jw})|) = \lg(|P(e^{jw})|) + \lg(|V(e^{jw})|) \quad (1.6)$$

де

$S(e^{jw})$ – спектр мови;

$P(e^{jw})$ – спектр сигнала збудження;

$V(e^{jw})$ – частотна характеристика голосового тракту

Якщо виконати над $\lg(|S(e^{jw})|)$ зворотне перетворення Фур'є, то отримаємо так званий кепстр сигналу.

Параметри голосового тракту змінюються в часі порівняно повільно (їх спектр знаходиться в області низьких частот - НЧ), тоді як сигнал збудження - швидкоосцілюруюча функція (її спектр зосереджений в області високих частот - ВЧ). Тому в кепстрі мовного сигналу ці складові поділяються і можуть бути закодовані окремо.

Структурна схема гомоморфного кодера мови приведена на рисунку 1.10.

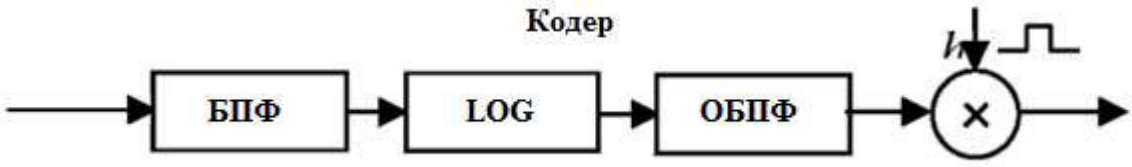


Рисунок 1.10 – Структурна схема гомоморфного код ера

При використанні гомоморфного вокодера при передачі мовного сигналу можна отримати швидкість коду близько 4,8 кбіт / с.

1.3.5 Вокодери з лінійним передбаченням

Вокодери на основі лінійного передбачення використовують таку ж модель мовоутворення, що й інші з розглянутих. Їх відмінною особливістю є метод визначення параметрів тракту. Лінійні передбачувані кодери (ЛПК), налагоджують голосовий тракт лінійним фільтром з безперервною імпульсної перехідною характеристикою, в якому кожне чергове значення

сигналу може бути отримано як лінійна комбінація деякого числа від його попередніх значень.

У ЛПК-вокодерах мовний сигнал ділиться на блоки довжиною близько 20 мс, для кожного з яких визначаються коефіцієнти передбачуваного фільтра.

Ці коефіцієнти квантуються і передаються декодеру. Потім мовний сигнал пропускається через фільтр, частотна характеристика якого зворотня частотній характеристиці голосового тракту.

На виході фільтра виходить помилка передбачення. Призначення передбачення - усунути кореляцію між сусідніми відліками сигналу.

У результаті набагато виразніше проявляється довготривала кореляція в сигналі, що дозволяє точніше визначити частоту основного тону і виділити ознаку голосний / приголосний звук.

Вокодери на основі лінійного передбачення зараз найбільш популярні, оскільки всі використовувані ними фільтровані моделі мовного тракту працюють дуже добре.

Отримані з їх допомогою швидкості кодів при непоганій якості мови складають до 2,4 кбіт / с.

1.4 Гибридні методи кодування мови

Гібридні, або комбіновані, методи кодування мови заповнюють розрив між кодерами форми сигналу, абсолютно не враховуючими його природи, і кодерами джерела, кодуючими, по суті, не сигнал, а параметри моделі.

Як зазначалося раніше, кодери форми сигналу забезпечують дуже добру якість мови при швидкостях кодування вище 16 кбіт/с, але взагалі не працюють при більш низьких швидкостях, тоді як вокодери забезпечують розбірливу мову при швидкостях кодування 4,8 кбіт/с і нижче, але не можуть дати доброї якості при будь-якій швидкості коду.

До них на зміну прийшли гібридні методи кодування, що працюють в тимчасовій області (тобто з сигналом, а не його спектром або іншими лінійними перетвореннями), засновані на аналізі сигналу через його синтез (так звані ABS-кодери).

Ці кодери так само, як і вокодер, використовують модель голосового тракту, але трохи іншим чином - для підбору сигналу збудження, що забезпечує найкращий збіг синтезованого сигналу на основі мовного сигналу з вихідним.

ABS-кодери були запропоновані порівняно недавно. В своєму первісному вигляді отримали назву MPE-кодерів (Multi-Pulse Excited – окодери з багатоімпульсним збудженням).

Пізніше були запропоновані більш досконалі RPE-кодери (Regular-Pulse Excited - кодери з регулярним імпульсним збудженням) і CELP-кодери (Codebook-Excited Linear Predictive – зі збудженням на основі кодових книг).

1.4.1 Кодери на основі метода «аналіз через синтез»

ABS-кодер працює таким чином. Закодований вхідний сигнал (вже в цифровій формі, у вигляді потоку відліків) розбивається на фрагменти довжиною порядку 20 мс, в межах яких властивості сигналу змінюються незначно.

Для кожного з цих фрагментів визначаються поточні параметри синтезуючого фільтра (аналога голосового тракту) і далі підбирається сигнал збудження, який, мінімізує помилку між вхідним і синтезованим сигналами.

Таким чином, назва методу Analysis-by-Synthesis полягає в тому, що кодер аналізує вхідну мову за допомогою синтезу безлічі наближень до неї.

Зрештою кодер передає декодеру інформацію, що представляє собою комбінацію поточних параметрів синтезу фільтра і сигналу збудження. Бажано, щоб цих даних було поменше. Декодер за цими параметрами відновлює закодовану мову, причому робить це так само, як це робив кодер в процесі аналізу через синтез.

Різниця між ABS-кодерами різного типу полягає в тому, що в кожному з них підбирається сигнал збудження синтезуючого фільтра u(n).

Теоретично на вхід синтезуючого фільтра потрібно подати нескінченно велику кількість різних сигналів збудження, щоб подивитися, який сигнал буде на його виході, і порівняти його з закодованим. Сигнал збудження, який дає мінімум зваженої помилки між оригіналом і синтезованої мовою, вибирається в якості результату кодування.

Саме ця замкнута схема визначення сигналу збудження (рис. 1.11) і забезпечує ABS-кодерам високу якість кодуючої мови при низьких швидкостях коду.

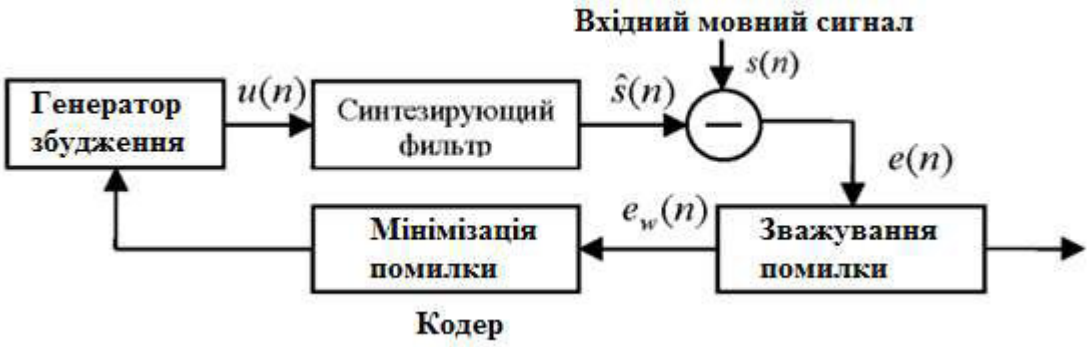


Рисунок 1.11 – ABS – кодер

У ABS - кодері полягає проблема великіої кількості обчислювальних операцій, необхідних для відбору найкращого сигналу збудження. Але для сьогоднішніх можливостей обчислювальної і мікропроцесорної техніки це цілком здійсненне завдання.

1.4.2 Кодери з постійними імпульсними збудженнями

Кодер Regular Pulse Excited, або RPE-кодер, використовує в якості сигналу збудження u(n) фіксований набір коротких імпульсів. Однак у цьому кодері імпульси розташовані регулярно на однакових відстанях один від одного, і кодеру необхідно визначити лише положення першого імпульсу і амплітуди всіх імпульсів.

Таким чином, декодеру потрібно передавати менше інформації про становище імпульсів, отже, в сигнал збудження можна включити їх більшу кількість і тим самим поліпшити наближення синтезованого сигналу до оригіналу.

Наприклад, якщо при швидкості коду 10 кбіт/с в MPE-кодері використовується чотирьохімпульсний сигнал збудження, то в RPE-кодері можна використовувати вже десятиімпульсний сигнал. При цьому істотно підвищується якість мови.

Метод регулярного імпульсного порушення RPE сьогодні широко застосовується, в тому числі в системі стільникового зв'язку GSM.

1.4.3 Кодери зі збудженням на основі кодових книг

Метод кодування RPE забезпечує добру якість закодованої мови при швидкостях коду близько 8 кбіт /с і вище, але починає сильно спотворювати сигнал при більш низьких швидкостях.

Справа в тому, що для опису необхідних параметрів сигналу збудження - тимчасового положення і амплітуд імпульсів - з необхідною точністю, просто не вистачає біт.

У зв'язку з цим був запропонований метод, що використовує в якості сигналу збудження не імпульсні послідовності, що задаються набором своїх параметрів, а бібліотеки (кодові книги) спеціальним чином підготовлені і записані в запам'ятовуючий пристрій сигналів збудження різної форми – Codebook Excited Linear Prediction (CELP) .

Схема формування сигналу збудження для CELP-кодера наведена на рис. 1.12.

Результатом кодування при цьому не є параметри імпульсів сигналу збудження, а індекс кодової книги, а також його амплітуда.

Метод кодування CELP заснований на лінійній авторегресійній моделі процесу формування та сприйняття мови і входить до групи методів аналізу

через синтез, що реалізують сучасні та ефективні алгоритми інформаційного стиснення мовних сигналів.

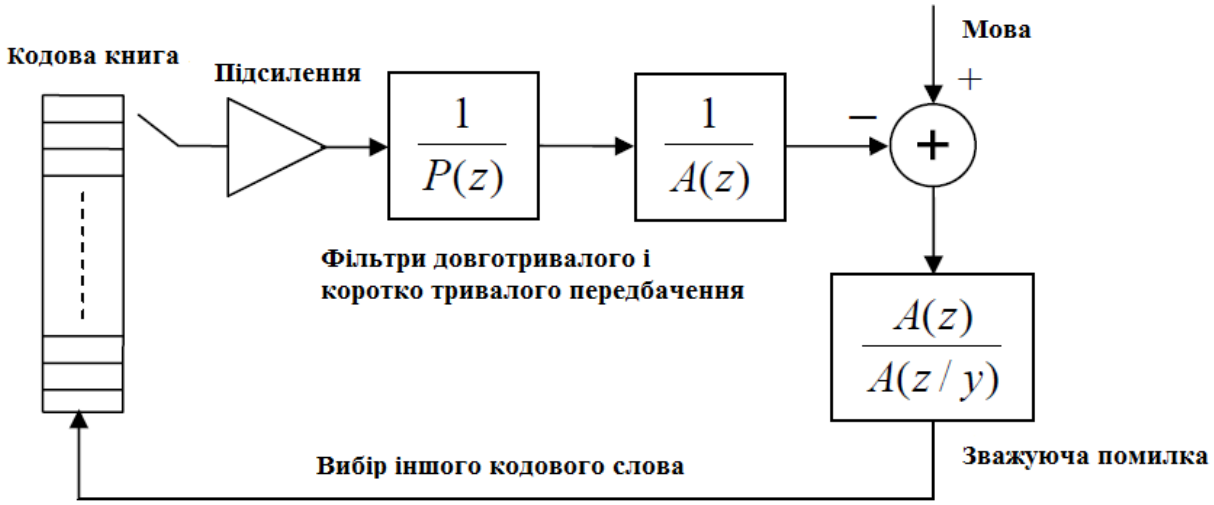


Рисунок 1.12 – Формування сигналу збудження CELP – кодера

Алгоритми даного класу займають проміжне положення між кодерами форми сигналу, в яких зберігається форма коливання мовного сигналу в процесі його дискретизації і квантування, і параметричними вокодерами, заснованими на процедурах оцінки та кодування невеликого числа параметрів мови, об'єднуючи переваги кожного з них.

Лінійна авторегресійна модель процесу формування мовних сигналів з локально постійними на інтервалах 10…30 мс параметрами отримала в даний час найбільшого поширення.

Для цієї моделі:

$$\lambda(h) = \sum_{m-1}^{M} a(m)\,\lambda(h-m) + x(h) \qquad (1.7)$$

де M — порядок моделі; v– послідовність відліків мовного сигналу; a(m) – коефіцієнти лінійного передбачення, які характеризують властивості голосового тракту; x(h)– породжуюча послідовність або сигнал збудження голосового тракту.

Авторегресійна модель мовного сигналу описує його з досить високим ступенем точності і дозволяє застосовувати розвинений математичний апарат

лінійного передбачення. При цьому забезпечується більш висока якість декодування мови, стійкість до вхідного акустичного шуму і помилок в каналі зв'язку в порівнянні з системами і з іншими принципами кодування.

У рамках даної моделі найбільш перспективними методами кодування вважаються методи «аналізу через синтез» з використанням багатоімпульсного збудження.

Новизна багатоімпульсного збудження полягає в тому, що в сигналі залишку лінійного передбачення вибираються такі його значення, які найбільш важливі для підвищення якості синтезованої мови. При цьому використовувана у процедурі аналізу через синтез схема кодування, крім обліку помилок квантування, включає критерії суб'єктивної оцінки якості мовного сигналу, що забезпечує природне звучання синтезованої мови.

При багатоімпульсному порушенні сигнал залишку лінійного передбачення представляється у вигляді послідовності імпульсів з нерівномірно розподіленими інтервалами і з різними амплітудами (близько 8-10 імпульсів за 10 мс).

Амплітуди і положення цих імпульсів визначаються на покадровій основі (кадр за кадром).

Основною перевагою багатоімпульсного збудження є те, що він визначається для будь-якого мовного сегмента і при цьому не потрібно знань ні про вокалізованность даного сегмента, ні про період основного тону.

Методи аналізу через синтез використовують синтезатор (декодер) мовного сигналу як складову частину пристрою кодування. При цьому завдання аналізу зводиться до процедури оцінки переданих в канал зв'язку параметрів мови, яка проводиться згідно з деяким критерієм неузгодженості між вихідним і декодованими сигналами.

Для обліку специфіки слухового сприйняття в якості критерію неузгодженості зазвичай використовується зважена за частотою квадратична помилка:

$$\varepsilon_w = \int_0^{F/2} |S(f) - S_q(f)|^2 \, w(f)df \quad (1.8)$$

де S(f) и $S_q(f)$ – перетворення Фур'є вихідного і синтезованого мовних сигналів; w(f) – вагова функція.

Беручи до уваги важливість для сприйняття мови не тільки формант, але і межформантних областей, для алгоритмів аналізу мови через синтез в якості еталонної була запропонована вагова функція такого вигляду:

$$w(z)=A(z)A^{-1}(z/\gamma) \quad (1.9)$$

де $A^{-1}(z)$ – передавальна характеристика синтезуючого фільтра; z/γ – параметр, який регулює енергію помилки або шум квантування.

Фактично при такому вікні зважування підкреслюється помилка у міжформантних областях і тим самим забезпечується більш рівномірне за частотою розподілення відношення потужності корисного сигналу до потужності помилки кодування.

В алгоритмах кодування з «аналізом через синтез» підвищення ефективності інформаційного ущільнення мовних сигналів виробляється, переважно, за рахунок скорочення надмірності послідовності x(h), яка здійснює збудження синтезуючого фільтра $A^{-1}(z)$ лінійного збудження, формуючого огинаючу сигналу, з коефіцієнтом передачі:

$$A^{-1}(z)=(1-\sum_{m-1}^{M} a(m)\, z^{-m})^{-1} \quad (1.10)$$

Для цієї мети застосовується також додатковий фільтр з характеристикою:

$$p^{-1}(z)=(1-g_p z^{-T})^{-1} \quad (1.11)$$

де g_p – коефіцієнт передбачення з затримкою на період основного тона T.

Фільтр виконує функції генератора квазіперіодичних коливань голосових зв'язок при вимові вокалізованих звуків.

Будучи однією з найпоширеніших, схема з лінійним передбаченням і збудженням від коду CELP є кращою схемою AbS-LPC для низьких швидкостей.

У CELP є лінійний фільтр зі змінюючими в часі параметрами для виділення грубої і точної спектральної інформації. Збудження виконується шляхом перебору всіх векторів із збуджуючої кодової книги. Векторна послідовність, що забезпечує мінімальну зважену помилку, вважається оптимальним збудженням.

Процедура AbS в CELP вимагає великих обчислювальних ресурсів, а основна кодова книга є результатом дуже великої дослідницької роботи.

Хоча CELP є складним методом, він здатний синтезувати мову з високою якістю навіть на низьких швидкостях. Варіант кодування CELP обраний для багатьох систем голосового зв'язку.

1.5 Параметри порівняння методів кодування мови

До найбільш важливих параметрів кодування відносяться:

1) Швидкість - це діапазон необхідних швидкостей передачі.

Для систем з більш низькою бітової швидкістю потрібна менша смуга частот, з цієї причини вони забезпечують більш високу ефективність використання спектру та потужності і в кінцевому підсумку призводять до стільникових систем радіозв'язку з підвищеною місткістю.

2) Якість - критерій,який використовується при порівнянні - наскільки добре звучить мова в ідеальних умовах - так звана чиста мова, без помилок передачі, тільки при кодуванні (зауважимо, однак, що в реальності ці ідеальні умови зустрічаються не часто, так як існує велика кількість типів фонових шумів, таких, як шум вулиці, шум робочого приміщення, шум у літаку та інше).

Слід зазначити, що якість є суб'єктивним результатом вимірювання та оцінювання.

3) Порогова ймовірність помилки на біт - більш високе значення порогової помилки веде до більш робастної структурі системи, а, отже, до

більш низьких вимог до відношення сигнал/шум і збільшення ємності мережі.

4) Затримки кодування системи передачі мови - це чинник, тісно пов'язаний з характеристикою якості системи. Затримка кодування складається з алгоритмічної затримки (буферизації мовлення для аналізу), обчислювальної затримки (часу, необхідного для обробки збережених мовних вибірок) і затримки в процесі передачі. Тільки перші дві відносяться до підсистеми кодування мови, хоча дуже часто схема кодування побудована так, що передача може початися навіть до того, як буде повністю оброблено інформацію в аналізованому фреймі. Для телефонних мереж, де ехо - головна проблема і воно має бути зведено до мінімуму, мала затримка відіграє велику роль.

Для рухомих систем і систем супутникового зв'язку широко застосовується процедура придушення еха, так як присутня відчутна затримка розповсюдження.

Однак у випадку з телефонними мережами, де затримка дуже мала, при використанні кодера з великими затримками будуть вимагатися надпотужні подавлювачі еха, що збільшать загальну вартість системи.

Інша проблема, затримки кодера - часто суб'єктивний дратівливий чинник.

Найкращі низькошвидкісні алгоритми вводять істотну затримку кодування в порівнянні зі стандартними ІКМ системами, розрахованими на швидкість передачі 64 кбіт / с.

У додатках типу телеконференцій може виникнути необхідність з'єднання декількох учасників через багатокрапковий пристрій управління, щоб зв'язати кожну людину з іншою.

Це вимагає декодування кожного потоку даних, підсумовування декодованих сигналів, а потім перекодування результуючого сигналу.

Цей процес подвоює затримку і, одночасно, зменшує якість мови з-за неодноразового кодування. Така система телемосту може допустити

максимальну затримку в одну сторону 100 мс, так як міст призведе до подвоєння затримки системи в одну сторону до 200 мс.

5) Складність і споживання енергії. У міру виникнення все більш «витончених» алгоритмів неминуче зростає складність обчислень.

Поява мікросхем цифрової обробки сигналів (DSP – Digital Signal Processor) і спеціалізованих BIC (ASIC – Application Specific Integrated Circuits) дало можливість значно знизити вартість обчислювальних операцій.

Проте показник складність / споживання енергії, а відповідно, і ціна - все це є основною проблемою, особливо для портативної апаратури.

Однією з основних технологій зниження споживання енергії та збільшення ефективності каналу, є цифрова інтерполяція мови (DSI - Digital Speech Interpolation). DSI використовує той факт, що мова, насправді, займає лише близько половини часу переговорів, і під час «простоїв» канал може бути використаний для інших цілей. Це дозволяє знижувати активність передавача, тобто економити енергію.

6) Обробка інших сигналів звукового діапазону. Канали зв'язку, спочатку призначені для передачі мовних сигналів, нерідко використовуються для передачі іншого роду сигналів, таких як сигнали від модему, факсу та ін. Статистичні параметри форми і частотний спектр таких сигналів зовсім відмінні від відповідних параметрів мови, тому алгоритм обробки повинен оперувати з обома видами сигналів.

При розробці алгоритму можливість перетворення немовних сигналів звукового діапазону часто відходить на другий план і розглядається на заключному етапі, що, безумовно, є помилкою, особливо при розробці кодерів для мереж загального користування.

Якість мови та швидкість передачі - два конфліктуючих фактора. Чим нижча швидкість мовного кодера, тобто більш високий ступінь компресії сигналу, тим більше страждає якість - в даному випадку ступінь достовірності.

Для систем, які працюють в телефонних мережах загального користування і подібних їм, основною вимогою є якість кодування (кодери форми). Однак для закритих мереж, таких, як приватні комерційні або військові системи зв'язку, вимоги до якості можуть бути знижені з метою підвищення швидкості.

Хоча зазвичай потрібно абсолютна якість, цим нехтують для іншої групи стандартів, мета яких - домогтися більш високого ступеня стиснення (гібридні кодери). Наприклад, в системах рухомого радіозв'язку висока ступінь стиснення при середній якості - це вирішальний фактор при виборі стандарту кодування.

Ця якість звичайно приймається в розрахунок як при добрих, так і при поганих умовах передачі. Виявляється, що перший фактор відносно просто виміряти щодо зменшення швидкості цифрового потоку. Визначити та оцінити спотворення важче.

Причина полягає в тому, що мовний сигнал сприймає не пристрій, а людина. На жаль, відсутність моделей сприйняття не дозволяє розробити об'єктивні методи оцінки спотворень мови.

Таблиця 1.1 – Усереднені характеристики стандартів стиснення мови

Стандарт	Мовний кодер	Швидкість, кбіт/с	Оцінка MOS	Затримка, мс
ITU – G.711	ІКМ (PCM)	64	4.3	0.125
ITU – G.711	АДІКМ (ADPCM)	32	4.0	0.125
ITU – G.726	VBR - (ADPCM)	16, 24, 32, 40	2.0, 3.2, 4.0, 4.2	0.125
ITU – G.728	LD - CELP	16	4.0	0.625
ITU – G.EFR	ACELP	13	4.0	20
FS1016	CELP	4.8	4.0	15
GSM Half – rate	VCELP	5.6	3.5	24.375

В таблиці 1.1 приведені MOS-оцінки для різних швидкостей передачі відповідно до стандартів на алгоритми кодування мови міжнародного телекомунікаційного союзу (ITU) і європейського інституту стандартів у галузі телекомунікацій (ETSI).

Представлені середні характеристики методів кодування, де VBR-ADPCM - кодер з АДІКМ і змінною швидкістю передачі, LD-CELP-вокодер з лінійним передбаченням, з кодовим збудженням і низькою затримкою,CS-ACELP - кодери з лінійним передбаченням, з алгебраїчним кодовою збудженням з сполученої структурою, VSELP - кодери зі збудженням векторної суми і лінійним передбаченням, ACELP-кодери з алгебраїчним кодовим збудженням і лінійним передбаченням, MP-MLQ - кодери з використанням багатоімпульсного збудження з квантуванням і алгоритму максимальної правдоподібності.

Принципова перевага перспективних кодерів полягає в тому,що вони можуть досягати гарної якості при значно нижчій швидкості передачі мовного сигналу (високим ступенем кодування).

1.6 Висноввок до першого розділу

Аналізуючи вищесказане можна зробити висновок:

1. Метод кодування CELP заснований на лінійній авторегресійній моделі процесу формування та сприйняття мови і входить до групи методів аналізу через синтез, що реалізують сучасні та ефективні алгоритми інформаційного стиснення мовних сигналів.

Авторегресійна модель мовного сигналу описує сигнал збудження голосового тракту з досить високою точністю і дозволяє застосовувати розвинений математичний апарат лінійного передбачення.

Її застосування забезпечує стійкість до вхідного акустичному шуму і помилок в каналі зв'язку в порівнянні з іншими принципами кодування.

2. Експериментально встановлено, що кодове збудження забезпечує найбільш високу якість кодування мовного сигналу, в тому числі і за наявності вхідних акустичних перешкод.

3.Будучи однією з найпоширеніших, схема з лінійним передбаченням і збудженням від коду CELP є кращою схемою AbS-LPC для низьких швидкостей..

В даний час спостерігається активний розвиток і впровадження нових засобів зв'язку і телекомунікацій.

Використання цифрового представлення даних дозволяє забезпечити надійність і економічність зв'язку. При цьому, воно призводить до багаторазового збільшення займаної полоси частот і, як наслідок цього, до багаторазового збільшення швидкості передачі інформаційних повідомлень.

Внаслідок цього, велике значення набуває вирішення проблеми мінімізації числа біт, необхідних для передачі сигналу, тобто проблема кодування мови.

Рішення проблеми лежить в області розробки ефективних методів цифрового перетворення (кодування).

1.7 Постановка задачі до другого розділу

На основі викладеного вище прийнято рішення, що найбільш відповідним алгоритмом кодування є кодування лінійного передбачення мультикодовим управлінням.

Метою даної кваліфікаційної роботи є підвищення якості мовного сигналу без збільшення швидкості передачі за рахунок реконфігурації мультисмугової кодової книги.

Для підвищення якості мовного сигналу без збільшення швидкості передачі необхідно вирішити наступні завдання:

1. Вдосконалити метод кодування лінійного передбачення мультикодовим управлінням.

2. Розбити частотну вісь на субсмуги.

3. Реконфігурувати структуру кодової книги на основі субсмугової перцептуальної ентропії.

4. Виконати математичне моделювання вдосконаленого методу з метою обгрунтування вибору методу кодування.

2 СПЕЦІАЛЬНА ЧАСТИНА

2.1 Банки цифрових фільтрів

Банк фільтрів (БФ) це цифровий пристрій, в якому вхідний сигнал, представлений послідовністю відліків, з допомогою М різних цифрових фільтрів розбивається на М різних канальних сигналів (для обробки деяким способом кожного з них), з яких за допомогою вихідних фільтрів і наступного підсумовування утворюється послідовність відліків вихідного сигналу.

Основну ідею побудови системи аналізу/синтезу сигналів з використанням банку фільтрів розкриває рис. 2.1.

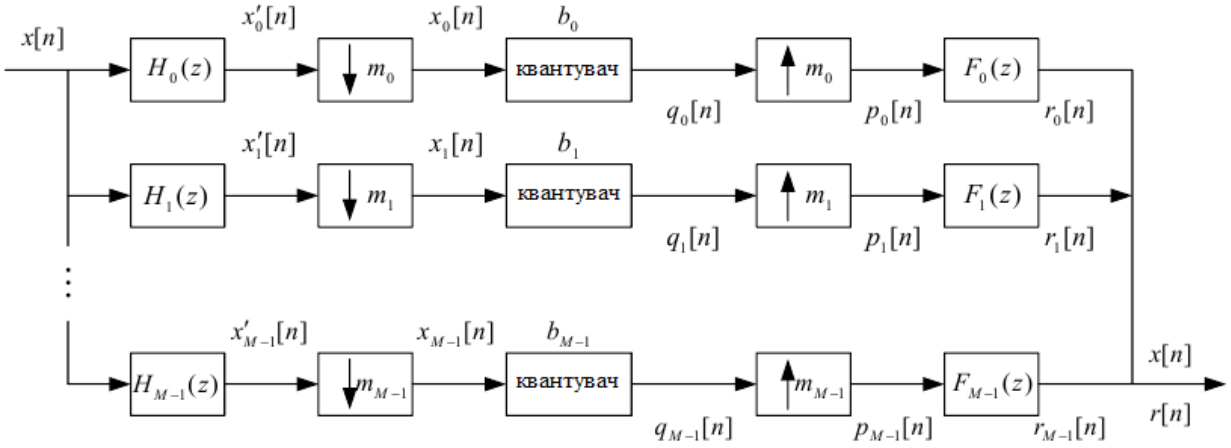


Рисунок 2.1 – Система аналізу/синтезу сигналів на основі банку фільтрів

Вихідний сигнал розбивається за допомогою фільтрів аналізу H_k (z), к=0,1…, М-1 на М субсмугових складових, які в ідеальному випадку в частотній області не перекриваються. Підібравши необхідним чином набір фільтрів синтезу F_k (z), к=0,1…, М-1, можна відновити вихідний сигнал з його субсмугових компонент.

Внаслідок обмеження ширини спектру сигналів на виході БФ можна зменшити частоту дискретизації субсмугових сигналів пропорційно зменшенню ширини спектру. Для зниження частоти дискретизації на стадії

аналізу і подальшого підвищення на стадії синтезу використовуються відповідно компресори та експандери частоти дискретизації.

У разі, коли коефіцієнт проріджування в кожному каналі дорівнює відношенню ширини спектру субсмугового сигналу до ширини спектру вихідного чи

$$\sum_{k=0}^{M-1} \frac{1}{m_k} = 1 \qquad (2.1)$$

говорять про систему аналізу/синтезу з повною децимацією. Таким чином,виходить М сигналів, що відображають поведінку вихідного сигналу в кожному частотному піддіапазоні, які представлені в сумі тією ж кількістю відліків, що й вихідний сигнал.

Кожен субсмуговий сигнал окремо може бути ефективно оброблений по деякому алгоритму b_k, к=0,1…, М-1.

Традиційно БФ поділяють на банки з рівносмуговими і нерівносмуговими каналами, ортогональні, біортогональні, двоканальні та багатоканальні і т.д. Кожен фільтр банку цифрових фільтрів утворює канал. Тому говорять про М-канальний банк фільтрів.

Сигнал в каналі називається субсмугою, звідси назва субсмугова фільтрація або субсмугове кодування. Рівносмугова декомпозиція має на увазі однаковий коефіцієнт децимації і однакову сумарну ширину смуги пропускання кожного каналу. У цей клас цифрових БФ входять також банки з багатокомпонентними фільтрами (мають більше однієї смуги пропускання).У разі нерівносмугових каналів коефіцієнти децимації різні і в загальному випадку можуть бути виражені у вигляді раціонального числа .

$$m_k = \frac{q_k}{p_k} \qquad (2.2)$$

У зв'язку з тим, що неможливо побудувати фільтр з ідеальною прямокутною АЧХ, доводиться миритися з наявністю у смугового фільтра перехідної зони і нерівномірності в смузі пропускання.

Ця обставина в поєднанні з ефектом «згортання» частот, що виникають при децимації, може призвести до значних спотворень при відновленні

сигналу. Однак існують методики розрахунку БФ, які дозволяють повністю (БФ з повним відновленням) або майже повністю, тобто з будь-якою наперед заданою точністю, відновити вихідний сигнал.

Дециматор (компресор частоти дискретизації) - пристрій, який здійснює децимацію (проріджування) сигналу в часі. Децимація - операція, що полягає у виключенні (пропуск) відліків вхідного сигналу, з порядковим номером, кратним коефіцієнту децимації.

Децимація в M раз позначається зазвичай як:

$$y(n) = (\downarrow M)x(n) = x(nM) \tag{2.3}$$

У частотній області це запишеться як:

$$Y\left(e^{j\omega}\right) = \frac{1}{M}\sum_{k=0}^{M-1} X\left(e^{\frac{i}{M}(\omega-2\pi k)}\right), \tag{2.4}$$

тобто спектр вихідного сигналу операції децимації містить M копій «розширеного» в M раз спектру вхідного сигналу, як це показано на рисунку 2.2.

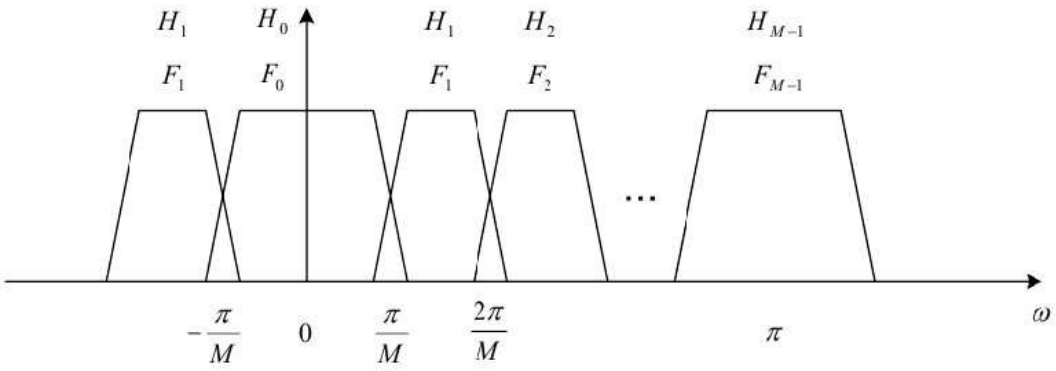


Рисунок 2.2 – Децимація сигналу в M раз

Як видно з рисунка 2.2, якщо сигнал не обмежений смугою частот

$$-\frac{\pi}{M} < \omega < \frac{\pi}{M} \tag{2.5}$$

то відбувається накладення спектрів копій, тобто алайзінг. Тому в банку фільтрів перед децимацією виконується НЧ-фільтрація. Сукупність фільтру і дециматора називається фільтром-дециматором.

Інтерполятор - пристрій, що виконує дії, зворотні децимації. Інтерполяція – операція, яка полягає у додаванні між відліками, чий

порядковий номер кратний певному числу, деякої константи (звичайно нуля). Інтерполяція в M раз зазвичай позначається як:

$$u(n) = \begin{cases} якщо\ \frac{n}{M}, тоді\ x\left(n/M\right) \\ інакше\ 0 \end{cases} \quad (2.6)$$

В частотній області це запишеться як:

$$U\left(e^{j\omega}\right) = X\left(e^{jM\omega}\right), \quad (2.7)$$

тобто спектр вихідного сигналу операції інтерполяції містить M копій «стисненого» в M раз спектру вхідного сигналу.

Ці копії повторюються через

$$\omega_k = \frac{2\pi k}{M} \quad (2.8)$$

Для їх усунення після інтерполятора ставиться НЧ-фільтр. Сукупність інтерполятора і фільтра називається фільтром-інтерполятором.

Еквівалентні схеми включення інтерполятора і дециматорів. Дві тотожності – дозволяють зменшити число обчислень в банку фільтрів. Це обумовлено тим, що частота дискретизації вхідного сигналу знижується за допомогою дециматора, а потім здійснюється процес фільтрації, таким чином загальна обчислювальна складність зменшується пропорційно коефіцієнту децимації.

Після закінчення обробки субсмугового сигналу в кожному каналі банку фільтрів частота дискретизації підвищується за допомогою інтерполятора. Дані рівності (еквівалентні схеми включення) широко застосовуються для виведення різних співвідношень і представлені на рисунку 2.3.

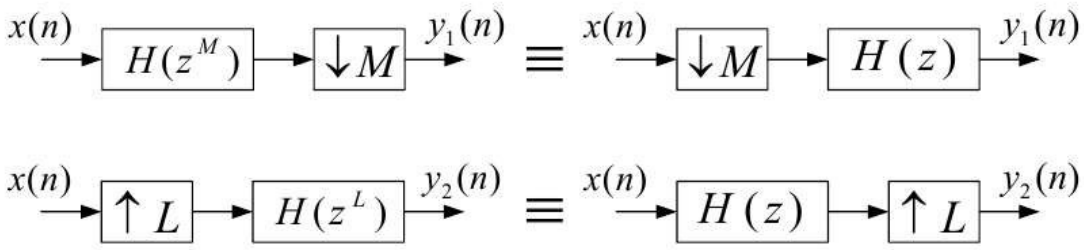


Рисунок 2.3 – Еквівалентні схеми включення фільтрів-дециматорів і фільтрів-інтерполяторів

Приведення коефіцієнтів децимації. Якщо банк фільтрів має нерівні коефіцієнти децимації, він може бути трансформований в банк фільтрів з рівними коефіцієнтами децимації, як описано нижче.

Нехай N = lcm(m_0, m_1, ... m_{M-1}). Канал з коефіцієнтом децимації/інтерполяції показаний на рисунку 2.4.

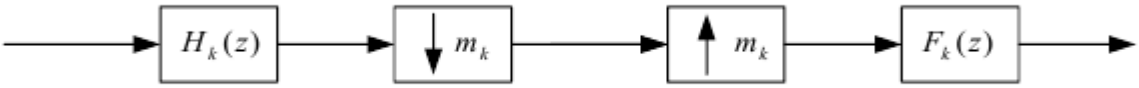


Рисунок 2.4 – К-й канал банка фільтрів

Уявімо систему з ланцюгом затримок з коефіцієнтом пере дискретизації

$$\frac{N}{m_k} \tag{2.9}$$

між стороною аналізу та синтезу, як показано на рисунку 2.5.

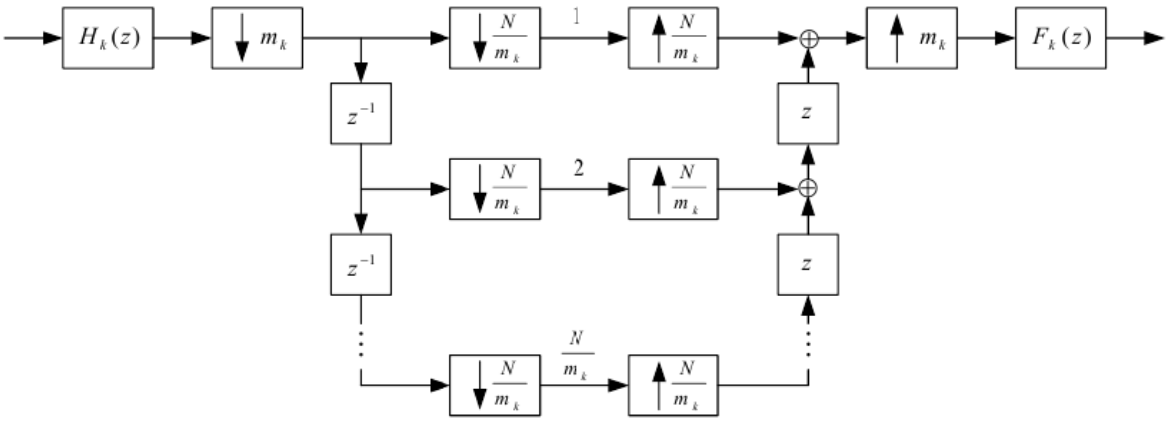


Рисунок 2.5 – К-й канал з ланцюгом затримок

Обидві системи, представлені на рисунку 2.4 і рисунку 2.5 ідентичні, так як система з ланцюгом затримок володіє перфективною реконструкцією і має нульову затримку (позитивна ступінь z використовується на синтезуючій стороні).

Перестановка затримок в представленій системі до m_k – дециматора і після m_k – інтерполятора приводить до системи, представленої на рис. 2.6.

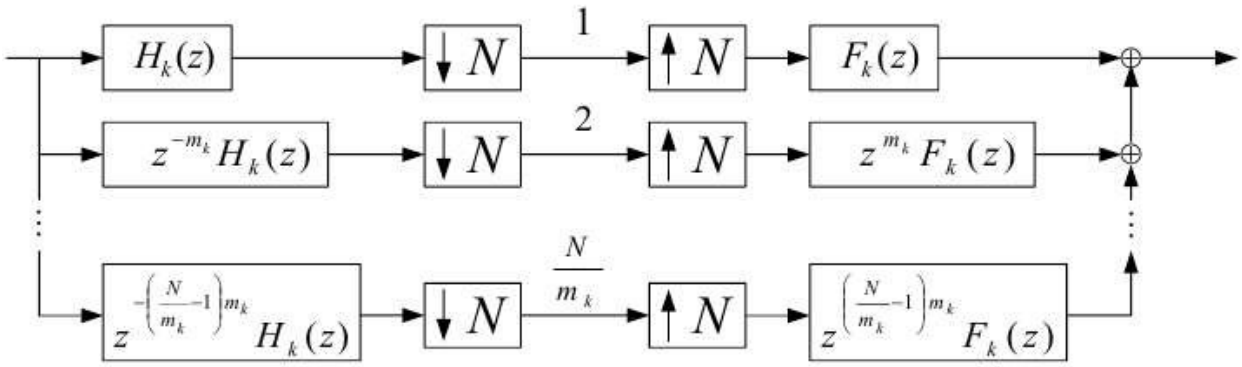


Рисунок 2.6 – Еквівалентна система к-го каналу

Фактично один канал з коефіцієнтом децимації m_k замінюється на N/m_k каналів з коефіцієнтів децимації N.

Подібним способом всі канали $0 \leq \kappa \leq M-1$ можуть бути перетворені до систем з коефіцієнтом децимації N. Використовуючи рівняння (2.1), максимальну кількість каналів можна представити наступним виразом

$$\frac{N}{m_0}+\frac{N}{m_1}+\cdots+\frac{N}{m_{M-1}}=N \tag{2.10}$$

Таким чином, банк фільтрів з нерівними коефіцієнтами децимації виражається як банк фільтрів з рівними коефіцієнтами децимації, що має N каналів.

Повне відновлення. Повне відновлення або перфективна реконструкція - властивість банку цифрових фільтрів, що полягає в тому, що сигнал, що пройшов через схему аналізу-синтезу ідентичний вхідному з точністю до затримки. Для цього фільтри синтезу повинні пригнічувати алайзінг і усувати амплітудні і фазові спотворення:

$$F_0(z)H_0(z)+F_1(z)H_1(z)=2z^{-1}, l\ mod\ 2\neq 0,$$

$$F_0(z)H_0(-z)+F_1(z)H_1(-z)=0,) \tag{2.11}$$

де F_k – синтезуючий фільтр к-го каналу, H_k – анілізуючий фільтр к-го каналу. Модуляційна матриця повинна задовольняти співвідношенню

$$F_m(z)H_m(z) = \begin{bmatrix} z^{-1} & 0 \\ 0 & z^{-1} \end{bmatrix} \quad (2.12)$$

Поняття параунітарної системи. Банк фільтрів є ортогональним (параунітарним), коли аналізують і синтезують фільтри і їх відповідно зміщені версії ортогональні один одному.

Нехай $x[n] = \begin{bmatrix} x_0(n) \\ \vdots \\ x_{N-1}(n) \end{bmatrix}$ буде вхідним вектором і $x[n] = \begin{bmatrix} y(n) \\ \vdots \\ y_{N-1}(n) \end{bmatrix}$ буде вихідним вектором з $N \times N$ параунітарною передавальною матрицею A(z).

Нехай $S_{xx}(e^{jw})$ буде N × N СПМ-матриця вхідного вектора x[n]. Зауважимо, що СПМ і-го вхідного компонента $x_i[n]$ є і-й елемент $S_{xx}(e^{jw})$, тому дисперсія складе

$$\int_{-\pi}^{\pi} (S_{xx}(w))_i \frac{dw}{2\pi} \quad (2.13)$$

Усереднена дисперсія вхідного сигналу буде

$$\frac{1}{N} = \sum_{i=0}^{N-1} \delta_i^2 = \frac{1}{N} \int_{-\pi}^{\pi} tr\left(S_{xx}(e^{jw})\right) \frac{dw}{2\pi}, \quad (2.14)$$

де tr(B) – слід матриці B.

СПМ вектора на виході визначається як

$$S_{yy}(e^{jw}) = A(e^{jw}) S_{xx}(e^{jw}) A^H(e^{jw}) \quad (2.15)$$

де A^H – матриця, ермітово транспонована до матриці A.

Усереднена дисперсія виходу складе

$$\frac{1}{N} \int_{-\pi}^{\pi} tr\left(S_{yy}(e^{jw})\right) \frac{dw}{2\pi} = \frac{1}{N} \int_{-\pi}^{\pi} tr\left(A(e^{jw}) S_{xx}(e^{jw}) A^H(e^{jw})\right) \frac{dw}{2\pi} \quad (2.16)$$

Так як tr(AB) = tr(BA), це спрощує вираз

$$\frac{1}{N} \int_{-\pi}^{\pi} tr\left(S_{xx}(e^{jw}) A^H(e^{jw}) A^H(e^{jw})\right) \frac{dw}{2\pi} \quad (2.17)$$

Так як матриця A(z) – параунітарна, $A^H A(e^{jw}) = I$, отже

$$\frac{1}{N} \int_{-\pi}^{\pi} tr\left(e^{jw}\right)\left(S_{xx}(e^{jw})\right) \frac{dw}{2\pi} \quad (2.18)$$

або усереднена дисперсія виходу дорівнює усередненій дисперсії входу.

Цей факт показує, що в параунітарній системі енергія зберігається.

Відомо, що для ортогонального банку фільтрів з рівними коефіцієнтами децимації поліфазної матриці аналізу та синтезу є параунітарними.

Поліфазна матриця, визначена для банків фільтрів з рівним коефіцієнтом децимації, може бути розширена до нерівних коефіцієнтів децимації,так як для схеми на рисунку 2.6 може бути сконструйована її еквівалентна $N \times N$ матриця.

Ці поліфазні матриці також параунітарні. Подібний ортогональний банк з нерівними коефіцієнтами децимації може бути побудований, використовуючи декілька ортогональних банків фільтрів з рівними коефіцієнтами децимації в деревовидній структурі.

2.1.1 ДПФ-модульований рівносмуговий банк фільтрів на основі поліфазної структури фільтрів дециматорів і інтерполяторів

При проектуванні банку цифрових фільтрів з рівносмуговими каналами використовується модель з повною модуляцією. Якщо банк заснований на одному НЧ-фільтрі прототипі з кінцевою імпульсною характеристикою (КІХ), то ширина смуги пропускання фільтра-прототипу визначає ширину кожного каналу.

Щоб такий фільтр міг виділити смугу, відповідну кожному каналу, необхідно зрушити спектр в область низьких частот при допомозі гетеродина (експоненціального модулятора), а потім здійснити НЧ-фільтрацію фільтром-прототипом. Після чого можна знизити частоту дискретизації рівносмугового сигналу без втрати інформації.

Зниження частоти дискретизації здійснює компресор,який видаляє M-1 відліків з кожної послідовності довжиною M. Максимальний коефіцієнт децимації дорівнює кількості каналів K, таким чином, даний банк фільтрів є максимально децимірованним. Синтез здійснюється в зворотній послідовності.

Спочатку збільшується частота дискретизації. У експандері між кожними двома відліками вставляються M-1 нульових відліків. Потім

здійснюється фільтрація субсмугових сигналів з наступною модуляцією з метою переміщення субсмуги до відповідного частотного діапазону, який вона займала у вихідному широкосмуговому сигналі. Підсумовування виходів всіх каналів синтезуючого банку фільтрів дає відновлений широкосмуговий сигнал (рисунок 2.7).

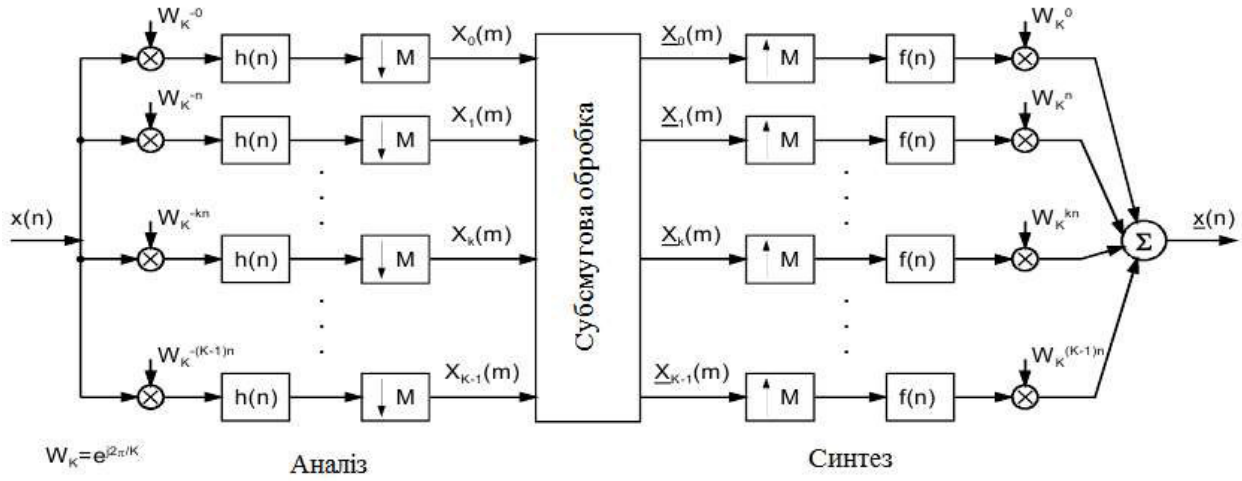


Рисунок 2.7 – Модель ДПФ-модульованого банку фільтрів, заснованого на повній модуляції

НЧ-фільтр прототип може бути спроектований стандартними методами, такими як синтез за допомогою зважують вікон, частотної вибірки і т.д. Частота зрізу фільтра-прототипу визначає кількість каналів та їх ширину, так як вся смуга може бути поділена на K рівних частин. Ступінь накладення субсмуг обмежується згідно з вимогами, що накладаються областю застосування конкретного банку фільтрів. Імпульсна і частотна характеристика фільтра для кожного каналу визначається наступним чином:

$$h_k(n) = h(n)W_K^{kn} = h(n)e^{j2\pi kn/K} \quad (2.19)$$

$$H_k\left(e^{jw}\right) = H\left(e^{j\left(\omega - \frac{2\pi k}{K}\right)}\right) \quad (2.20)$$

Система характеризується рівномірним розміщенням смуг з кроком

$$\omega_к = \frac{2\pi к}{K} \quad к = 0,1,\ldots К\text{-}1 \quad (2.21)$$

Безпосередня реалізація такої схеми банку фільтрів є вкрай неефективною. У кожному каналі при великій частоті дискретизації здійснюється згортка з імпульсною характеристикою фільтра-прототипу, що призводить до значного збільшення обчислювальних витрат, які можна знизити шляхом зниження частоти дискретизації. Ключем для побудови ефективної структури є полі фазна декомпозиція фільтра-прототипу. Вона ґрунтується на розбитті, децимації,групуванні коефіцієнтів фільтра на підгрупи, звані поліфазними фільтрами.

$$H(z) = \sum_{k=0}^{M-1} z^{-1} G_k(z^M) = G_0(z^M) + z^{-1} G_1(z^M) + \cdots + z^{-(M-1)} G_{M-1}(z^M)) \quad (2.22)$$

$$G_k(z) = h(k) + h(k+M)z^{-1} + h(k+2M)z^{-2} + \cdots$$

Таке групування поліфазних фільтрів може бути поділено між каналами (рис. 2.8 і 2.9).

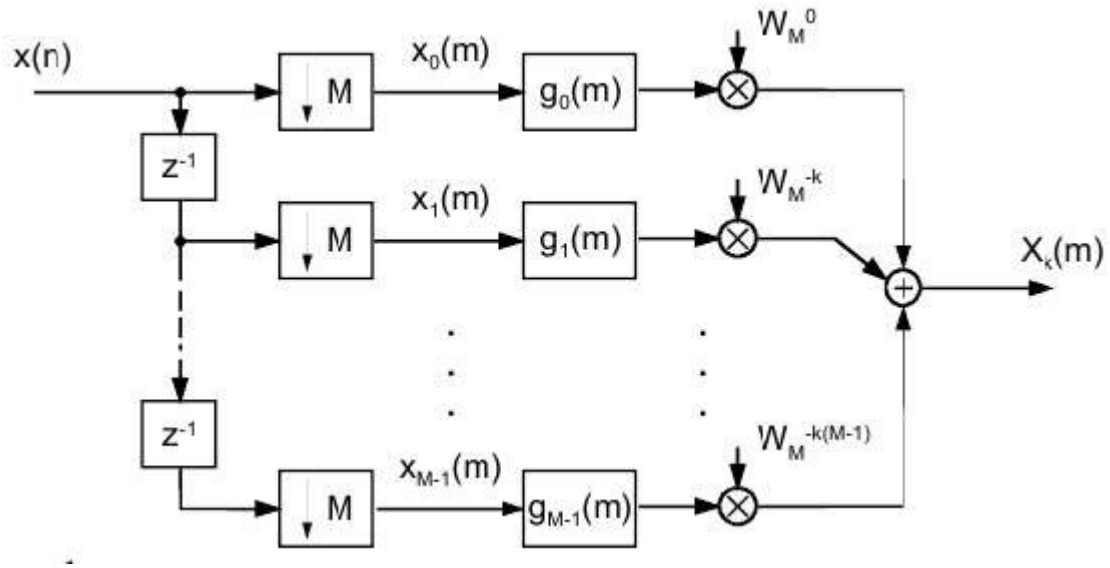


Рисунок 2.8 – Поліфазна структура каналу максимально дециміpованого банку аналізу

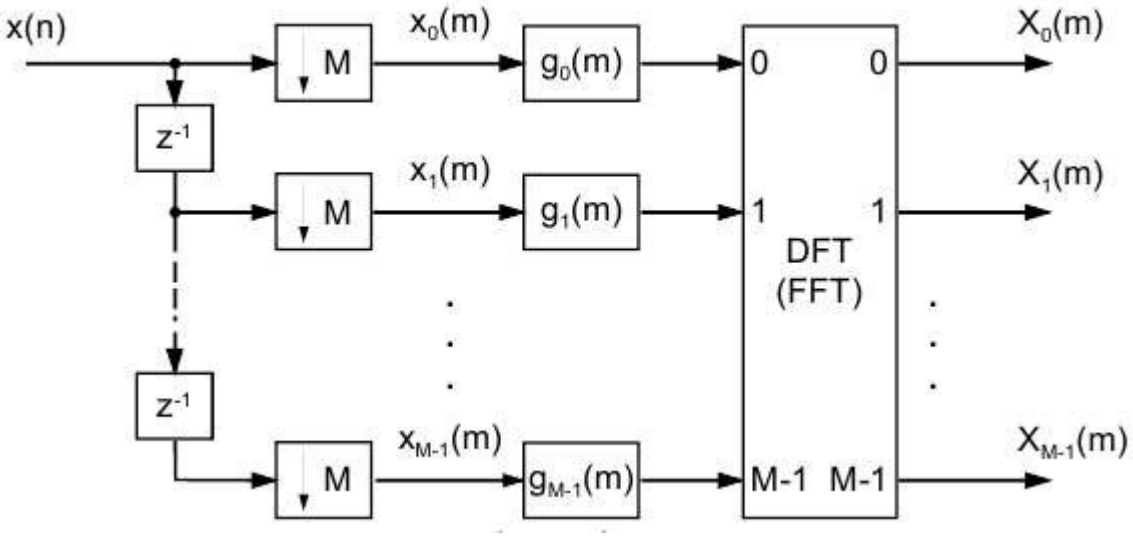


Рисунок 2.9 – ДПФ-модульований банк фільтрів аналізу з рівносмуговими каналами на основі поліфазної структури фільтрів децима торів

2.1.2 ДПФ-модульований нерівносмуговий банк фільтрів

У ряді випадків розбиття вихідної смуги сигналу на субсмуги рівної ширини не завжди є необхідним. Прикладом цього може служити перцептуальне перетворення, де система повинна відображати нерівномірне розкладання на критичні частотні субсмуги згідно особливостям людського сприйняття акустичної інформації. Описаний вище ДПФ-модульований банк фільтрів з рівносмуговими каналами можна перетворити в банк з нерівносмуговими каналами розкладанням субсмуг за допомогою фазового фільтра (ланки) :

$$z^{-1} \Rightarrow A(z) = H_{AP}(z) = \frac{z^{-1} - a}{1 - az^{-1}} \qquad (2.23)$$

Каскадне розташування фазових ланок стає причиною зміщення частотних компонент сигналу. Фазовий ланка не впливає на амплітуди частотних компонент сигналу, а лише змінює їх фазу, при цьому фазова характеристика даного ланки є нелінійної функцією частоти:

$$\left|A(e^{jw})\right| = \left|A(e^{j\varphi(\omega)})\right| = 1 \quad (2.23)$$

(2.24)

$$\varphi(\omega) = -\omega - 2\arctan\frac{|a|\sin(\omega - a)}{1 - |a|\cos(\omega - a)}$$

Характер нелінійності фазової характеристики визначається параметром фільтра a. Фільтр є стабільним, при $|a| < 1$. Чим ближче значення модуля параметра до 1, тим сильніше виражена нелінійність фазової характеристики фільтра.

Для параметра $a = 0$ затримка однакова для всіх частотних складових сигналу. Компоненти вхідного сигналу з різними частотами, що пройшли через фазовий фільтр, затримуються неоднаково, так як групова затримка фільтра також є нелінійною функцією від пульсації:

$$\tau(\omega) = -\frac{d\varphi(\omega)}{d\omega} = \frac{-(1-|a|^2)}{(1-|a|^2)+2|a|(1-\cos(\omega-a))} \quad (2.25)$$

Послідовність, що пройшла через ряд таких фільтрів не є оригінальною затриманою послідовністю. Компоненти сигналу,залежні від своєї пульсації, переміщуються повільніше або швидше.

Таким чином, сигнал, що пройшов через послідовність фазових ланок,є деформованою в частотній області версією вихідного вхідного сигналу. Ще більш гнучкого розкладання на субсмуги можна досягти,застосовуючи фазові фільтри вищих порядків R > 1:

$$H_{AP_R}(z) = j^{a_0} \prod_{r=1}^{R} \frac{z^{-1} - a_r}{1 - a_r z^{-1}} \quad (2.26)$$

У таких фільтрах фазова функція є більш еластичною, що дозволяє здійснювати більш вільну деформацію субсмуг в банку фільтрів рисунок 2.10 – 2.13.

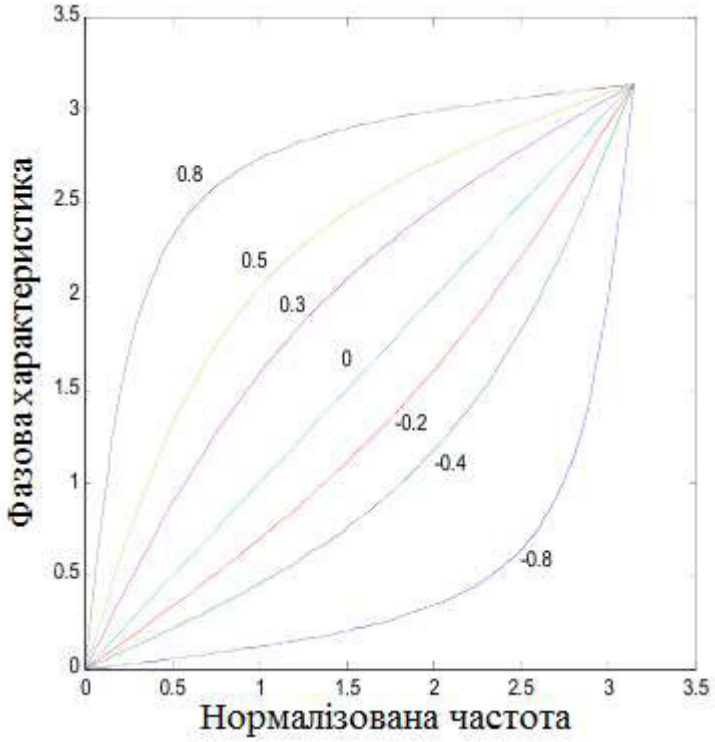


Рисунок 2.10 – Фазова характеристика фазової ланки першого порядку для різних коефіцієнтів фільтра

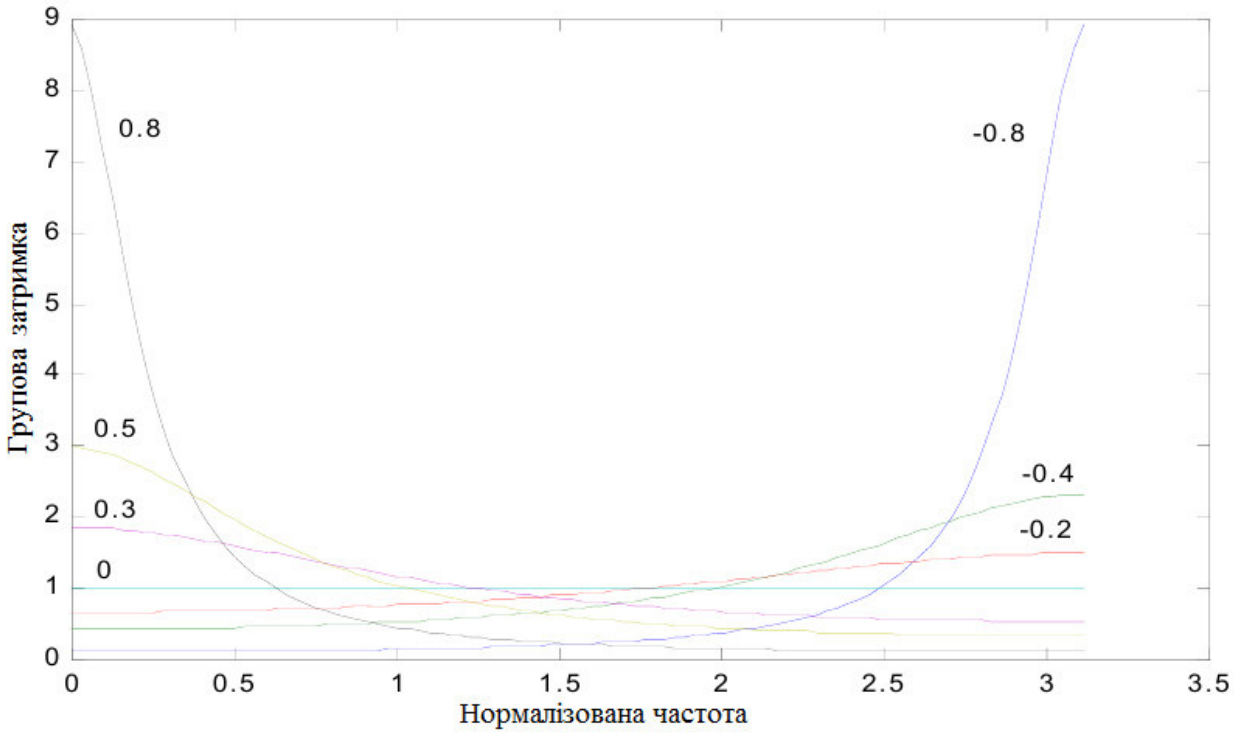


Рисунок 2.11 – Групова затримка фазового фільтру для різних значень коефіцієнтів

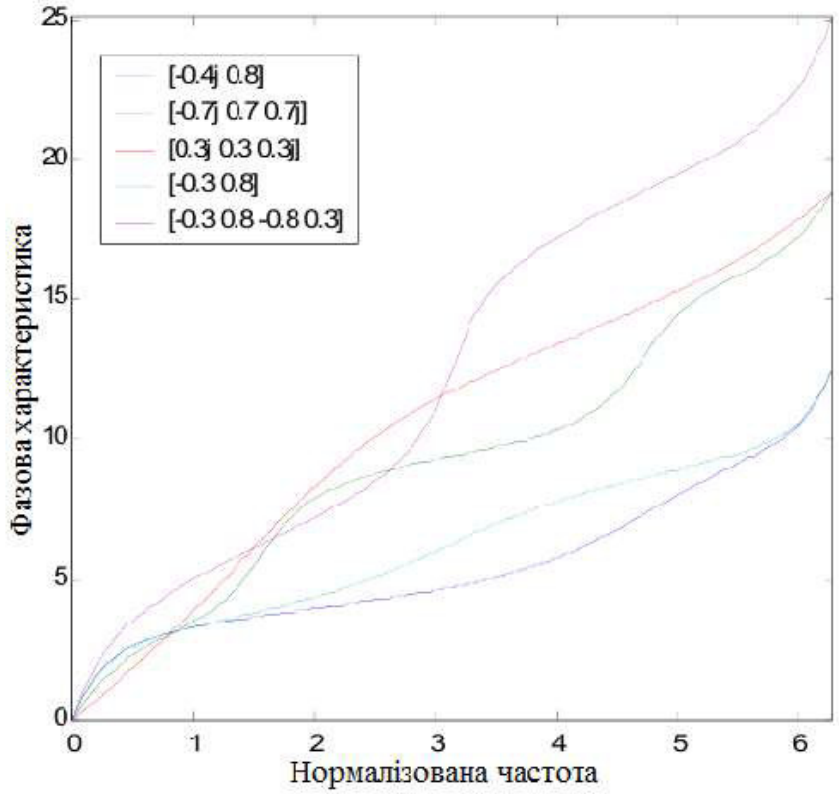


Рисунок 2.12 – Фазова характеристика фазових фільтрів вищих порядків для різних коефіцієнтів при багаторазовому перекритті

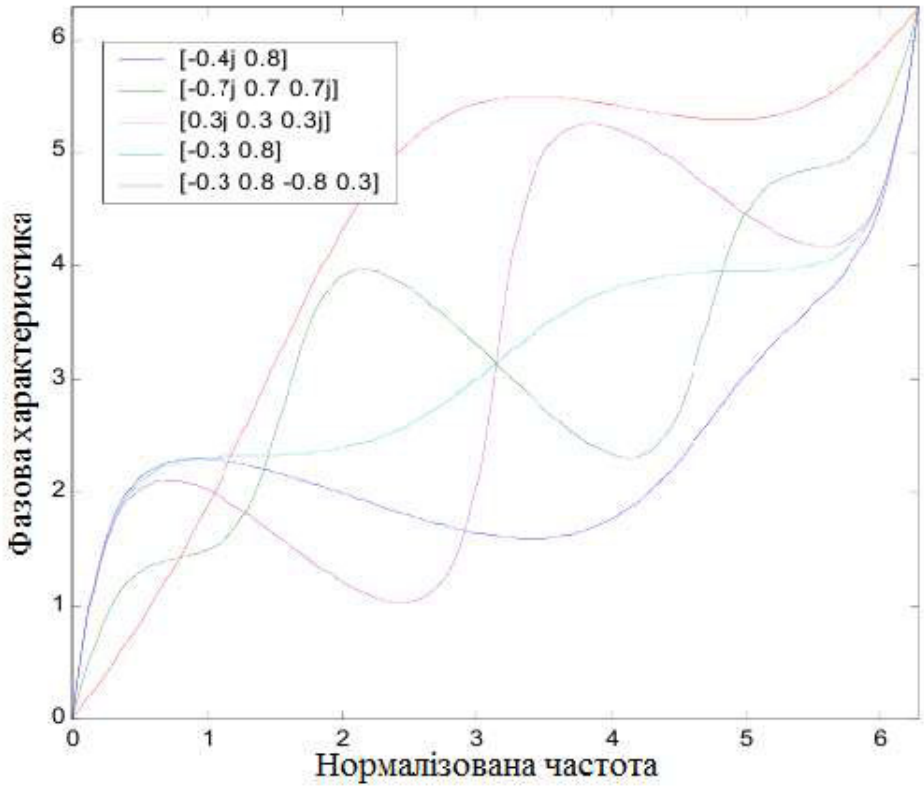


Рисунок 2.13 – Фазова характеристика фазових фільтрів вищих порядків для різних коефіцієнтів при одноразовому перекритті

У додатках кодування мовних сигналів ДПФ-модульовані банки не застосовуються. Найчастіше використовуються банки з косинусної модуляцією. Застосовувана в них модуляція реалізується косинусоїдальними функціями, внаслідок чого вхідний мовний сигнал після аналізу залишається дійсним (зберігається його структура).

Інші області застосування косинусних банків подібні ДПФ-модульованим. Часто вживаним косинусний-модульований банк з нерівносмуговими каналами є декомпозиція на субсмуги, згідно зі шкалою Барків, які узгоджені з психоакустичною моделлю слуху людини.

2.1.3 Приклад банку ДПФ-модульованих поліфазних фільтрів

Банк синтезований на основі фільтра прототипу з кінцевою імпульсною характеристикою. Число коефіцієнтів фільтра N = 640.

Кількість каналів в банку 64, тобто коефіцієнт децимації (у аналізуючому банку фільтрів) і інтерполяції (у синтезучому банку фільтрів) дорівнює 64.

При частоті дискретизації f_s = 16кГц виходить, що частота дискретизації в каналі f_s / M = 250Гц, смуга пропускання каналу становить 125Гц. На рисунку 2.14 наведена амплітудно-частотна характеристика ДПФ-модульованого поліфазного банку фільтрів аналізу-синтезу.

З рисунку 2.14 видно, що амплітудних спотворень в реконструйований сигнал (сигнал на виході синтезуючого банку фільтрів) не вноситься.

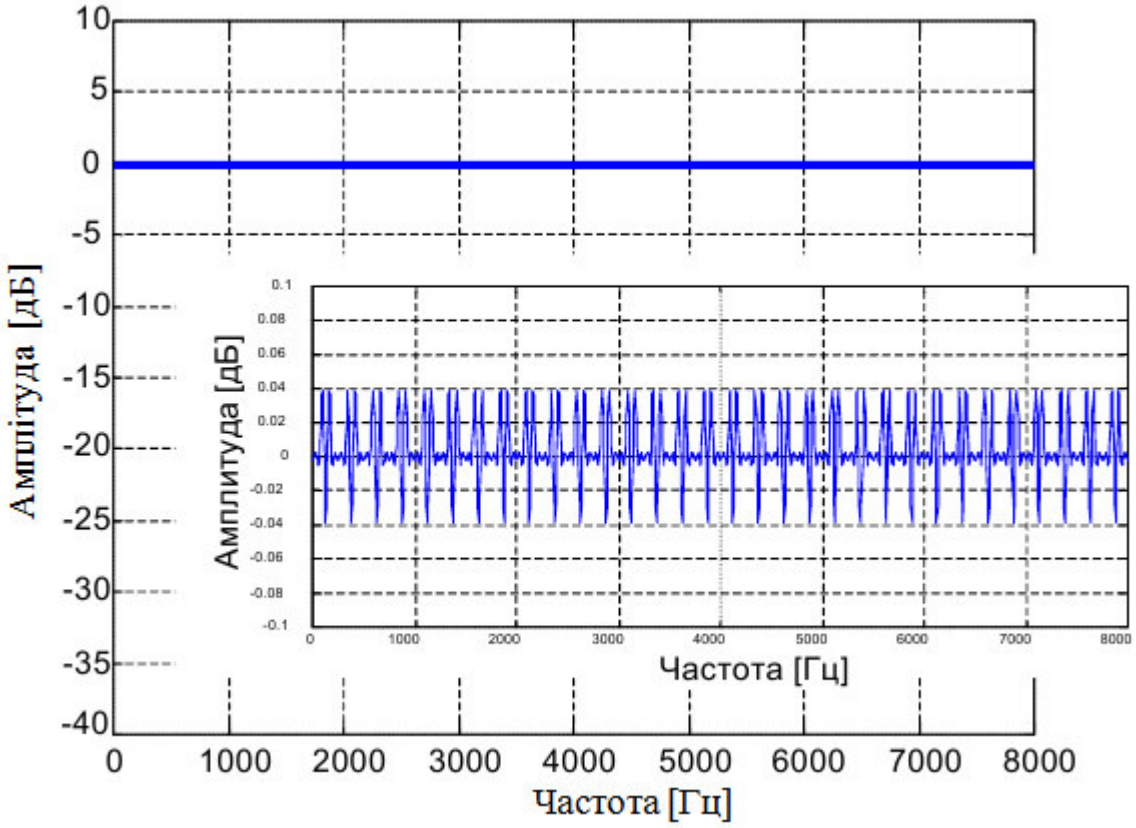


Рисунок 2.14 – Амплітудно-частотна характеристика поліфазного банку фільтрів аналізу – синтезу

2.2 Косинусно-модульовані банки фільтрів

2.2.1 Проектування косинусно-модульованого банку фільтрів

У М-канальному косинусно-модульованому банку фільтрів, фільтри аналізу та синтезу в загальному випадку описуються наступними виразами:

$$h_k(n) = 2h(n)\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)n + \varepsilon_k\right]$$

(2.27)

$$f_k(n) = 2h(n)\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)n + \gamma_k\right]$$

для к = 0,…,М -1, де h(n) і f(n) – НЧ КІХ-фільтри-прототипи з частотою зрізу $\pi/_{2M}$. Вибір фаз ε_k і γ_k може бути здійснений декількома способами. Для забезпечення перфективної реконструкції фази можна обчислити за виразами, представленим нижче:

$$\varepsilon_k = -\frac{\pi}{M}\left(k+\frac{1}{2}\right)\frac{\alpha}{2}$$

(2.28)

$$\gamma_k = +\frac{\pi}{M}\left(k+\frac{1}{2}\right)\frac{\alpha}{2}$$

У банку параунітарних фільтрів приймають

$$\varepsilon_k = -\frac{\pi}{M}\left(k+\frac{1}{2}\right)\frac{L+M-1}{2}+\frac{\pi}{2}$$

(2.29)

$$\gamma_k = -\frac{\pi}{M}\left(k+\frac{1}{2}\right)\frac{L+M-1}{2}-\frac{\pi}{2}$$

і, тим самим, використовують фільтр-прототип) h (n) = f (n) з довжиною імпульсної характеристики L = 2Mm, як на стороні аналізу, так і на стороні синтезу.

Підставляючи вирази (2.29) в (2.27), отримаємо

$$h_k(n) = 2h(n)\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{N-1}{2}\right)+(-1)^k\ \frac{\pi}{4}\right]$$

(2.30)

$$f_k(n) = 2h(n)\cos\left[\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(n-\frac{N-1}{2}\right)-(-1)^k\ \frac{\pi}{4}\right]$$

Швидка реалізація косинусної модуляції може бути здійснена шляхом використання ШПФ при незначній модифікації передавальних функцій фільтрів-прототипів:

$$H_k(z) = a_k b_k U(z) + a_k^*\ b_k^* U_k^*\ (z)$$

(2.31)

$$F_k(z) = a_k^* b_k V(z) + a_k\ b_k^* V_k^*\ (z)$$

де

$$U_k(z) = H\left(zW^{k+\frac{1}{2}}\right)$$

(2.32)

$$V_k(z) = F\left(zW^{k+\frac{1}{2}}\right)$$

і

$$a_k = e^{j^{(-1)^k \frac{\pi}{4}}}$$

$$b_k = W^{-\frac{1}{2}(k+\frac{1}{2})} \tag{2.33}$$

Модуляція реалізується так, щоб досягти найменшого перекриття між сусідніми субсмугами. Перфективна реконструкція може бути досягнута відповідним проектуванням фільтра-прототипу. Ключем для цього є запис поліфазної передавальної матриці для банку аналізу:

$$E(z) = C\begin{bmatrix} g_0(-z^2) \\ z^{-1} g_1(-z^2) \end{bmatrix} \tag{2.34}$$

де

$$g_0\ (z) = diag[G_0\ (z), G_1(z), \dots G_{M-1}\ (z)]$$

$$g_1\ (z) = diag[G_M\ (z), G_{M+1}(z), \dots G_{2M-1}\ (z)] \tag{2.35}$$

і

$$c_{kl} = 2\cos(\frac{\pi}{M}\left(k+\frac{1}{2}\right)\left(l-\frac{N-1}{2}\right) + (-1)^k\ \frac{\pi}{4}) \tag{2.36}$$

2.2.2 Косинусно-модульовані банки фільтрів з квазиповним відновленням

Існує ряд випадків, коли вимоги до перфективної реконструкції не є критичними. Часто це пов'язано з тим, що помилка реконструкції банку фільтрів може бути значно нижче, або порівнянна з помилкою модифікація субсмугових сигналів, оброблюваними будь-яким алгоритмом.

Більшість відомих методик проектування прототипу для косинусно-модульованих банків з квазіповним відновленням засновані на рівності

$$T\left(e^{j\omega}\right) = \sum_{k=0}^{M-1} F_k\left(e^{j\omega}\right) H_k\left(e^{j\omega}\right) = e^{j\omega(N-1)} \sum_{k=0}^{2M-1} \left| e^{j(\omega - k\pi/M)} \right|^2 \tag{2.37}$$

характеризує обмеження на лінійну фазо-частотну характеристику і зменшення ефекту «алайзінга». Обмеженням є лінійна фазочастотна характеристика фільтра-прототипу і частота зрізу π/M або

$$\left|T\left(e^{j\omega}\right)\right| \approx 0 \tag{2.38}$$

Таким чином, фільтр-прототип повинен бути спроектований так,щоб амплітудно-частотна характеристика $T\left(e^{j\omega}\right)$ була в наближенні плоскої (з мінімальними пульсаціями) для всіх частот ω або $\left|T\left(e^{j\omega}\right)\right| \approx 1$.

Запропоновано багато методів вирішення цього завдання з різними цільовими функціями і відповідними процедурами оптимізації коефіцієнтів. Наприклад, при пошуку коефіцієнтів фільтра-прототипу може використовуватися мінімізація наступної цільової функції:

$$\phi = max\left\{\left|H(e^{j\omega})\right| + \left|H(e^{j\left(\omega-{}^{\pi}/_{M}\right)})\right|^{2}\right\}$$

$$\omega \epsilon \left(0,{}^{\pi}/_{M}\right) \tag{2.39}$$

2.3 ДПФ з нерівномірною частотною роздільністю (WDFT)

2.3.1. Основні принципи WDFT

ДПФ з нерівномірним розкладанням коефіцієнтів Фур'є по частотної осі (WDFT) дозволяє отримати нерівномірне z-перетворення кінцевої послідовності, використовуючи фазовий фільтр A(z). WDFT послідовності x[n] з N точок визначається за наступною формулою:

$$X(z_k) = \sum_{n=0}^{N-1} x\,[n] z_k^{-n}\,, k = 0, \dots, N-1 \tag{2.40}$$

де z_k – зображення рівновіддалених точок на одиничному колі в z-площині.

Найпростіший варіант WDFT заснований на фазовій ланці першого порядку. Умовою стабільності є a <1. Залежно від знака a, розтягується низькочастотний ($a > 0$) або високочастотний (a <0) діапазон, таким

чином, що частина, що залишилася на одиничному колі стає стислою. Формально це може бути виражено:

$$\omega = \omega + 2\arctan\left(\frac{a\sin\omega}{1-a\cos\omega}\right) \tag{2.41}$$

тобто здійснюється білінійне перетворення z-площині в нову викривлену z-площину (рисунок 2.15).

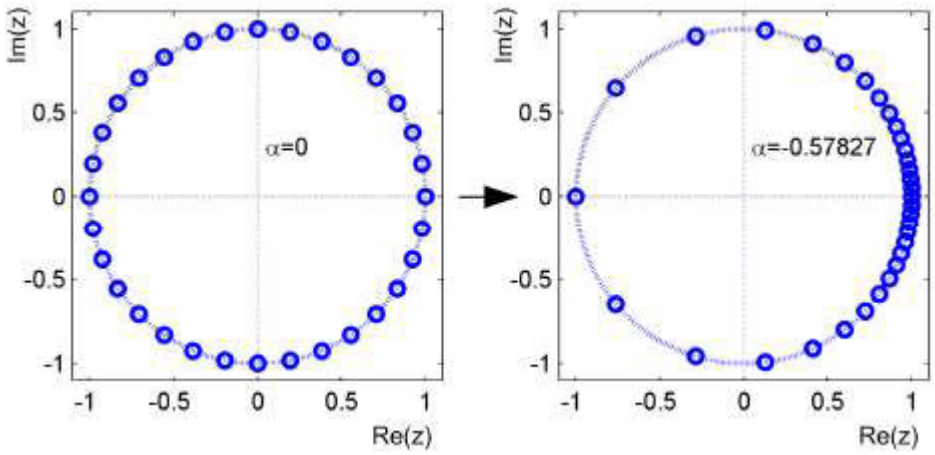


Рисунок 2.15 – Розміщення частотних відліків ДПФ (ліворуч) і WDFT (праворуч)

WDFT успадковує деякі властивості ДПФ. Одним з найважливіших є сполучена симетрія для дійсних даних:

$$X[N-K] = X^*[k] \tag{2.42}$$

Так як будь-який напрямок в просторі комплексних векторів,визначається базисним вектором, відповідає конкретному частотному діапазону, то WDFT можна представити як максимально децимірований банк фільтрів. При цьому, k-й рядок WDFT матриці може розглядатися як КІХ-фільтр з передавальною функцією, яка визначається за наступним виразом:

$$H_k(z) = \sum_{n=-\infty}^{\infty} A(z_k)z^{-1}, k = 0 \ldots N-1 \tag{2.43}$$

де A(z) – фазова ланка першого порядку;

$H_к(z)$ – смуговий фільтр з центральною частотою

$$\omega_k = 2tan^{-1}(\frac{1+a}{1-a}\tan(\frac{\omega_k}{2}))$$

(2.44)

$$\omega_k = angle(z_k)$$

Алгоритм, по ефективності порівнянний з ШПФ, не може бути побудований для WDFT через асиметричність WDFT-матриці. Тим не менш,прямий алгоритм роботи з комплексної матрицею може бути значно оптимізований.

Даний метод добре підходить для обробки зображень, де всі дані надходять одночасно. У обробці аудіосигналів, де відліки поступають послідовно один за іншим, може бути використана навіть пряма реалізація перетворення.

2.3.2 Зворотне перетворення WDFT

Уразі перцептуального деформування осі частот матриці Вандермонда, до класу яких відноситься і матриця D WDFT, володіють властивостями, що роблять неможливим отримання точної зворотної матриці. А саме, матриця WDFT може розглядатися як сингулярна так, як між деякими її рядками існують майже лінійні залежності. цей факт представляється як дуже мале значення її детермінанта $detD = \prod_{i<j}(z_i^{-1} - z_j^{-1})$.

Іншими словами, WDFT-матриця є погано обумовленою. Це означає, що будь-який чисельний алгоритм звернення матриці, застосований до подібної матриці є дуже чутливим до малих змін даних, якими зазвичай нехтують. Посилення помилки може бути оцінено, використовуючи число обумовленості матриці (власне значення матриці).

Як показано на рис. 2.16, число обумовленості перетвореної матриці залежить від її розмірності і величини коефіцієнта деформації і має дуже великі значення навіть при малорозмірних перетвореннях,які злегка відрізняються від ДПФ.

Погана обумовленість є невід'ємною властивістю матриць Вандермонда, пов'язаних з реальними проблемами. Єдиним винятком є випадок відсутності деформування частотної осі, коли WDFT стає звичайним ДПФ.

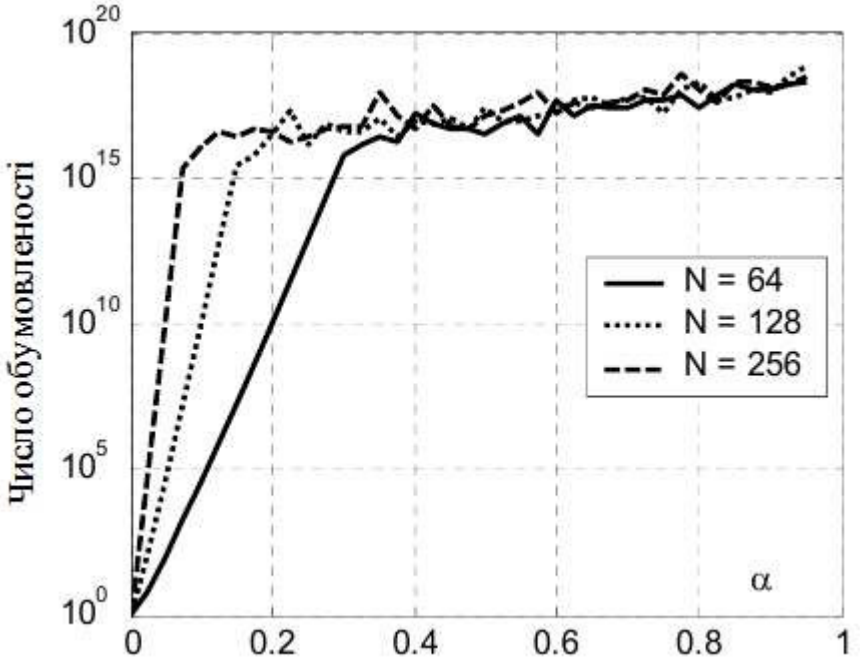


Рисунок 2.16 – Залежність числа обумовленості від коефіцієнта деформації для різних розмірів WDFT

Існує ряд методів для апроксимації звернень погано обумовлених матриць. Дані методи використовують розкладання посингулярними числами матриці, яке визначається як

$$D = U\Sigma V^H = \sum_{i=1}^{N} u_i \sigma_i v_i^H \tag{2.45}$$

де

$$U = [u_1, u_2, \dots u_N] \tag{2.46}$$

$$V = [v_1, v_2, \dots v_N]$$

матриці з ортогональними стовпцями $U^H U = V^H V = I_N$ і

$$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_N) \tag{2.47}$$

є діагональною матрицею складається з сингулярних чисел,відсортованих за зменшенням $\sigma_1 \geq \cdots \geq \sigma_N \geq 0$. Стовпці матриць U і V, лівий і правий, – сингулярні вектори.

Число обумовленості матриці D повного рангу є відношення $cond\,(D) = {}^{\sigma_{max}}/_{\sigma_{min}}$, де σ_{max} і σ_{min} – *н*найбільше і найменше сингулярні числа.

Ступінь неповноти матриці D (наявність деяких лінійних залежностей між її стовпцями) проявляється в існуванні майже нульових сингулярних чисел σ_i.

Компоненти SVD можуть дати багато корисних пояснень поганої обумовленості матриць. Також, вони можуть бути використані для формування псевдообернених матриць

$$\check{D} = \acute{\sum}_{i=1}^{N} f_i \frac{1}{\sigma_i} v_i u_i^H \qquad (2.48)$$

Символи f_i позначають так звані коефіцієнти ослаблення фільтра і повинні бути всі рівні одиниці для отримання точного звернення матриці. Теорія регуляризації рекомендує виключати вплив малих сигнулярних чисел, послаблюючи їх внесок у (2.48).

Це реалізується шляхом установки відповідних значень f_i. Так як сингулярні вектори з найбільшими індексами мають яскраво виражений коливальний характер, то ця дія змінює спектральний склад даних. Основна відмінність між відомими методами регуляризації полягає у виборі коефіцієнтів ослаблення фільтра. У найпростішому підході усіченого SVD сума (2.48) обмежується звичайним відкиданням термів мають сингулярні числа менше певного порогу

$$f_i = \begin{cases} 1 & \sigma_i > \lambda\,, \\ 0 & в\ іншому\ випадку \end{cases} \qquad (2.49)$$

Менш радикальним є демфіроване SVD, в якому коефіцієнти ослаблення фільтра поступово змінюються у бік нуля, забезпечуючи більш згладжене відсікання.

$$f_i = \frac{\sigma_i}{\sigma_i + \lambda} \tag{2.50}$$

В обох випадках, параметр регуляризації λ повинен бути вибраний з обережністю. Рішення має бути стабільним при спектральних змінах, обмежених певним мінімумом.

2.3.3 Оцінка помилки реконструкції WDFT-синтезу

Використовуючи матричну форму запису, вектор спотворення сигналу можна визначити як різниця між оригінальним і реконструйованим сигналами

$$d = x - \check{x} = (\mathrm{I} - \check{\mathrm{D}}\mathrm{D})x \tag{2.51}$$

Відповідна міра в частотній області може бути виражена як середня СПМ спотворення сигналу

$$S_{dd}(\omega) = \frac{1}{N} E\{|e(\omega)^H d|^2\} = \frac{1}{N} e(\omega)^H R_{dd}\, e(\omega)^H \tag{2.52}$$

де

$$e(\omega) = [1 \;\; e^{-j\omega} \;\; e^{-j\omega 2} \;\; e^{-j\omega(N-1)}]^T \tag{2.53}$$

є вектором ДПФ синусоїд R_{dd} позначає коваріацйну матрицю сигналу спотворення. Нехай $Q = \mathrm{I} - \check{\mathrm{D}}\mathrm{D}$, тоді

$$R_{dd} = Q R_{xx} Q^H \tag{2.54}$$

де R_{xx} – коваріаційна матриця оброблюваного сигналу.

Очевидно,що спектральне спотворення (2.52), яке визначається як абсолютна помилка,залежить від характеристик вхідного сигналу і якості апроксимації зворотного WDFT.

Теоретично, для ненульового R_{xx}, можливо точне зворотне перетворення, коли всі елементи дорівнюють нулю. Так як псевдообернена матриця розраховується методом розкладання за сингулярними числами матриці

$$Q = \mathrm{I} - \check{\mathrm{D}}\mathrm{D} = \mathrm{I} - \sum_{i=1}^{N} \mathrm{f_i}\ \mathrm{v_i}\ \mathrm{v_i^H} \tag{2.54}$$

відповідні налаштування коефіцієнтів фільтра можуть бути вельми корисні.

На жаль, у випадку перцептуальної деформації частотної осі,вибір параметрів регуляризації має мінімальний вплив на рівень спектральних спотворень. Було визначено, що навіть нестабільна апроксимація зворотного WDFT забезпечує відносно високу помилку реконструкції і остання посилюється при подальшій стабілізації.

Єдиний шлях мінімізації спектрального спотворення – це модифікація матриці перетворення таким чином, щоб кількість майже нульових сингулярних чисел було зменшено. Це ідентично зменшенню ексцентричності SVD-еліпсоїда, який є відображенням одиничної сфери в N-розмірному просторі.

На рисунку 2.17 показана СПМ пофарбованого гаусовського шуму і відповідне спотворення спектра. Можна помітити, що рівень спотворень на заданій частоті залежить від відстані між сусідніми WDFT коефіцієнтами.

Сигнал повністю відновлюється тільки в точках перетворення і спектральні спотворення особливо помітні у розтягнутих частотних діапазонах, в той час як в стислих діапазонах помилка синтезу має прийнятний рівень.

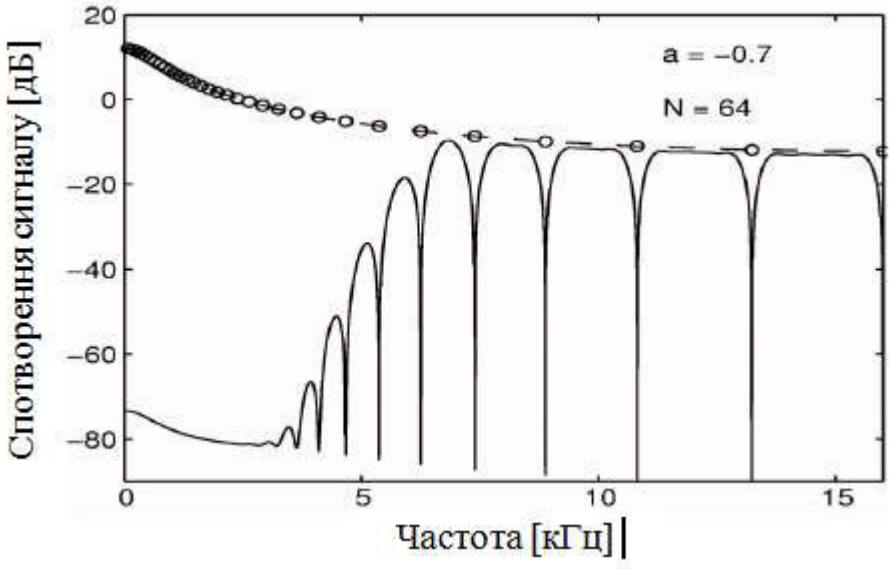


Рисунок 2.17 – СПМ оригінального сигналу і відповідна помилка синтезу

2.3.4 Формування надлишкового базису WDFT

k-й рядок WDFT матриці D є комплексним синусоїдальним вектором

$$s_k = \left[1 \; z_k^{-1} \; z_k^{-2} \; \dots \; z_k^{-N+1}\right] \tag{2.55}$$

Теоретично, для конкретних точок перетворення рядка WDFT матриці лінійно незалежні. Тому, синусоїдальні вектори WDFT формують неортогональний базис $S = \{s_0, s_1, s_2, \dots s_{N-1}\}$ для простору комплексних векторів N , так що будь - який вхідний вектор може бути виражений як їх лінійна комбінація.

Якщо ми послабимо обмеження на те, що вектори мають бути незалежними, то можна побудувати надлишковий не ортогональний базис. Незважаючи на те, що такий базис є надлишковим, він може внести значний внесок у корекцію помилки реконструкції, а саме, при відповідному виборі базису векторів, ми можемо модифікувати розподіл сингулярних чисел для відповідної матриці перетворення. Крім того, так як ми припускаємо надлишковий базис, нова WDFT матриця не є квадратною і кількість рядків M> N збільшується.

Узагальнена інверсія прямокутної матриці може бути легко знайдена, використовуючи SVD-процедуру, аналогічну для квадратної матриці. З практичної точки зору, існують як мінімум дві умови для побудови надлишкового векторного базису WDFT.

По-перше, необхідно зберегти регулярність точок перетворення. По-друге, розмір нового синусоїдального базису повинен бути по можливості мінімальним для забезпечення мінімізації обчислювальної складності.

Новий надлишковий базис повинен бути складений з M> N векторів,відповідно КІХ-фільтрам, чиї центральні частоти розташовані регулярно по всьому одиничному колі. Це може бути реалізовано через перевизначення точок перетворення:

$$z_k = e^{j\pi k/M}, k = 0 \dots M-1 \tag{2.56}$$

Якщо припустити від'ємне значення параметра всепропускаючого фільтра, то максимальне кутова відстань між новими точками z-перетворення складе

$$\Delta\omega_{max} = \pi - 2tan^{-1}\left(\frac{1+a}{1-a}\tan\left(\frac{\pi - \frac{2\pi}{M}}{2}\right)\right) \qquad (2.57)$$

Для отримання того ж частотного дозволу у ВЧ-діапазоні, що і у звичайного ДПФ, відстань між точками перетворення не повинно бути більше ніж 2π / N. Підставляючи $\Delta\omega_{max} = 2\pi/N$ в (2.44) і вирішуючи щодо M,виходить, що

$$M = M_{opt} = 2\pi[\pi - 2tan^{-1}(\frac{1+a}{1-a}\tan(\frac{\pi-2\pi/N}{2}))]^{-1} \qquad (2.58)$$

2.3.5 Апроксимація психоакустичних шкал

Перший крок при використанні WDFT в психоакустичної моделі - проектування відповідного всепропускаючого перетворення. Частотні коефіцієнти z-перетворення повинні бути представлені рівномірно в перцептуальній області. Показано, що всепропускаюча ланка першого порядку досить добре апроксимує перцептуальну шкалу барків і ERB, при цьому значення коефіцієнта всепропускаючого фільтра для заданої частоти дискретизації визначається за виразами (2.59) і(2.60), відповідно:

$$a_{\mathrm{Bark}} = 0.1957 - 1.048[\tfrac{2}{\pi}\arctan(0.07212\tfrac{f_s}{1000})]^{\frac{1}{2}} \qquad (2.59)$$

$$a_{\mathrm{ERB}} = 0.7164[\tfrac{2}{\pi}\arctan(0.09669\tfrac{f_s}{1000})]^{\frac{1}{2}} + 0.08667 \qquad (2.60)$$

Для випадку f_s= 16kHz коефіцієнт $a_{\mathrm{Bark}} = 57827.\ 0.$ На рис. 2.18 представлена трансформація (деформування) частотної осі відповідно до шкали Барків, ERB і MEL для f_s= 16kHz .

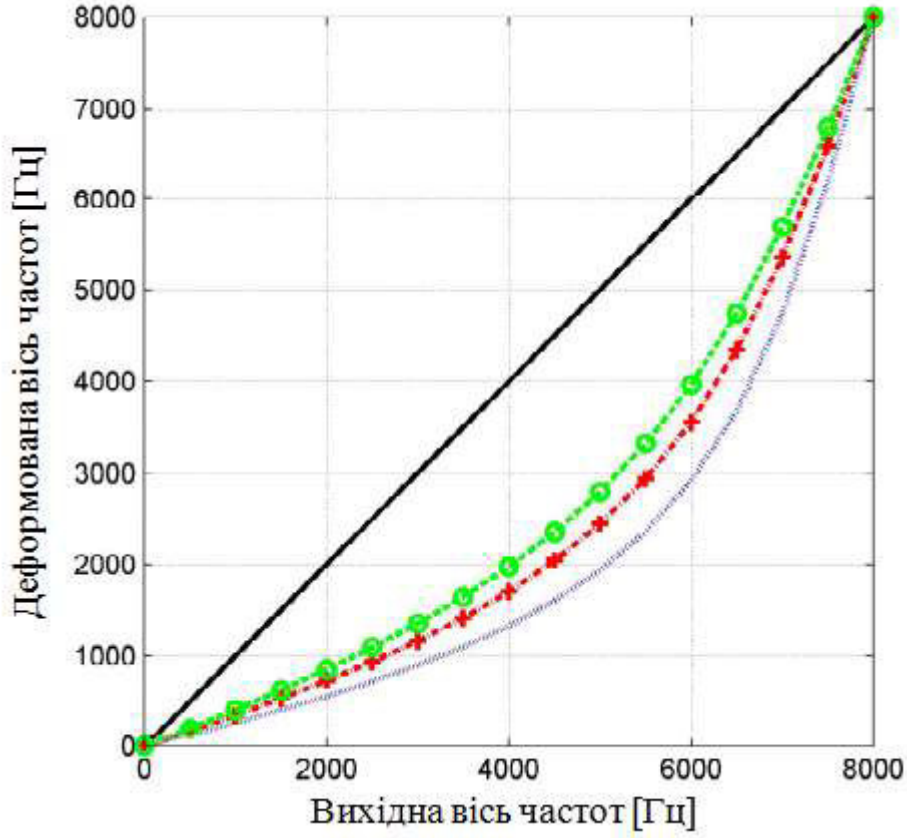


Рисунок 2.18 – Деформація частотної осі відповідно до перцептуальних шкал

Так як ДПФ має рівномірну частотну роздільність, у той час як ширина критичних частотних смуг суворо змінюється з їх місцем розташування на частотній шкалі, то різна кількість коефіцієнтів перетворення асоціюється з конкретною критичною частотною смугою. У частині А таблиці 2.1 кількість коефіцієнтів у групах варіюється від 3 до 38, в той час як для WDFT тієї ж розмірності не віддає переваги жодній зі смуг, всі коефіцієнти перетворення розподілені практично рівномірно (частина Б).

WDFT в його оригінальній формі не зберігає енергію сигналу в відповідних частинах на одиничному колі до і після всепропускаючого перетворення. Так як викривлення призводить до розтягування одного частотного діапазону при одночасному стисканні іншого, воно повинно супроводжуватися масштабуванням WDFT-амплітуд з метою корекції рівнів енергії в кожній критичній частотній смузі.

Розглянемо як приклад широкосмуговий аудіо сигнал, який містить голосове повідомлення (перші 3.5с) і музичний фрагмент з вокальним супроводом (рис. 2.19 (зверху)). Побудуємо спектрограми даного сигналу за допомогою ДПФ і WDFT (рис. 2.19 центр і низ,відповідно). Як видно з рисунку 2.19 WDFT-рішення дає досить чіткий результат, оцінка енергії більш точна. Представлені факти свідчать про те, що WDFT може дійсно бути базисом для психоакустичної моделі.

Таблиця 2.1– Відображення коефіцієнтів ДПФ і WDFT на критичні частотні смуги (Барки)

		Частина А розмір ДПФ=512, Fs=16 кГц				Частина Б розмір WDFT=512, Fs=16кГц		
Критична полоса		Діапазон коефіцієнтів	К-сть	Діапазон частот, Гц		Діапазон коефіцієнтів	К-сть	Діапазон частот, Гц
1		1 – 3	3	31 - 94		1 - 12	12	8 – 100
2		4 – 6	3	125 - 188		13 - 24	12	109 – 202
3		7 – 9	3	219 - 281		25 - 36	12	210 – 305
4		10 – 13	4	313 - 406		37 - 48	12	314 – 412
5		14 – 16	3	438 - 500		49 - 60	12	421 – 523
6		17 – 20	4	531 - 625		61 - 73	13	533 – 650
7		21 – 24	4	656 - 750		74 - 85	12	660 – 776
8		25 – 29	5	781 - 906		86 - 97	12	787 – 912
9		30 – 34	5	938 - 1063		98 - 110	13	923 – 1073
10		35 – 40	6	1094 - 1250		111 - 123	13	1086 - 1254
11		41 – 46	6	1281 - 1438		124 - 135	12	1269 - 1443
12		47 – 54	8	1469 - 1688		136 - 148	13	1460 - 1680
13		55 – 62	8	1719 - 1938		149 - 161	13	1700 - 1961
14		63 – 73	11	1969 - 2281		162 - 174	13	1985 - 2302
15		74 – 86	13	2313 - 2688		175 - 186	12	2331 - 2690
16		87 – 102	16	2719 - 3188		187 - 198	12	2726 - 3174
17		103 – 122	20	3219 - 3813		199 - 210	12	3220 - 3792
18		123 – 145	23	3844 - 4531		211 - 221	11	3851 - 4513
19		146 – 173	28	4563 - 5406		222 - 231	10	4588 - 5328
20		174 – 205	32	5438 - 6406		232 - 242	11	5419 - 6412
21		206 – 243	38	6438 - 7594		243 - 252	10	6520 - 7533
22		244 – 256	13	7625 - 8000		253 - 256	4	7650 - 8000

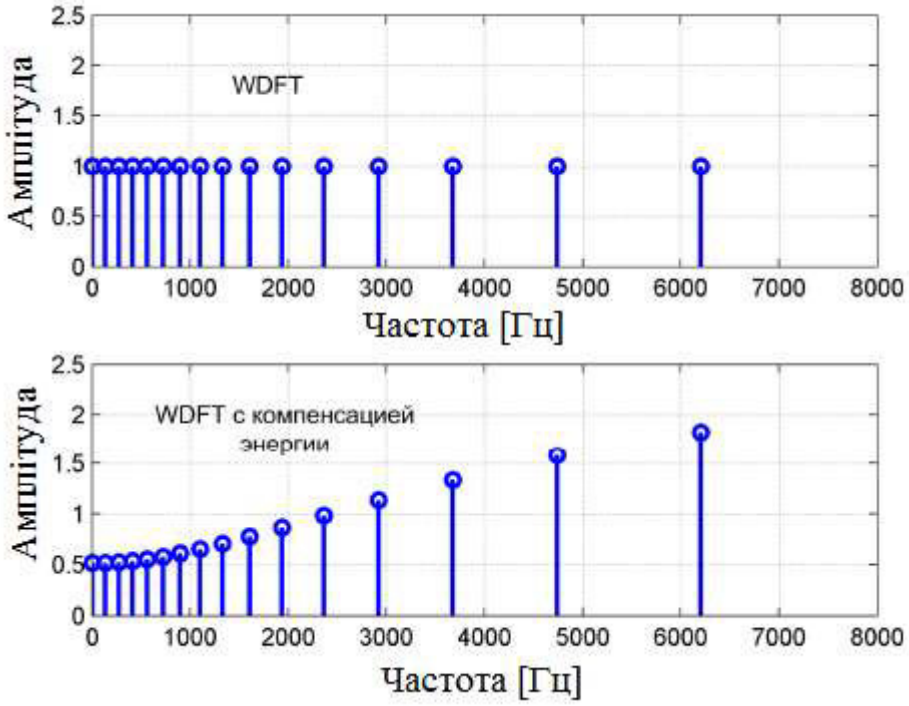


Рисунок 2. 19 – Компенсація енергії WDFT одиничного імпульсу

2.4 Побудова моделі кодування мови із застосуванням банків цифрових фільтрів і WDFT

2.4.1 Опис узагальненої моделі кодування мови

Алгоритм CELP (Code Excited Linear Prediction), заснований на рекомендації FS1016, побудований на моделі кодування з використанням процедури аналізу-через-синтез, лінійного передбачення і векторного квантування. Для моделювання короткочасного спектра мовного сигналу (формантної структури) використовується фільтр лінійного передбачення.

Для формування сигналу збудження використовуються адаптивна і стохастична кодові книги. Обчислювальна складність алгоритму визначається процедурами пошуку оптимальних векторів збудження по двом кодовим книгам, рис. 2.20.

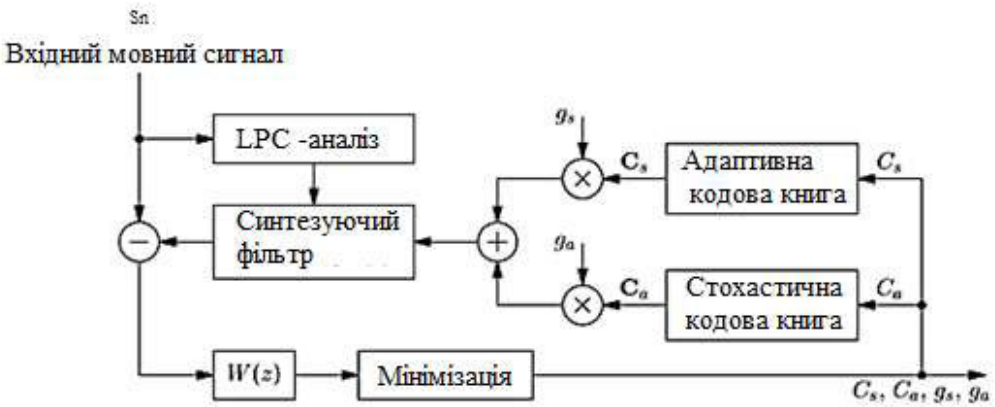


Рисунок 2.20 – Схема кодування алгоритму CELP

CELP аналіз складається з трьох основних процедур:

- короткочасне лінійне передбачення,
- довготривалий пошук по адаптивної кодової книзі,
- пошук по стохастичній кодовій книзі

У CELP-кодері FS-1016 (рис. 2.21) мовний сигнал S_n поділяється на кадри. У кожному кадрі з використанням алгоритму лінійного передбачення (LPC) визначаються параметри синтезуючого фільтра 1/A(z), після чого методом аналізу через синтез знаходяться параметри сигналу збудження, які мінімізують зважений сигнал помилки.

Сигнал збудження представляється наборами індексів (C_s, C_a векторів (C_s, C_a), витягають із стохастичної і адаптивної кодових книг, а також наборами відповідних їм коефіцієнтів посилення (g_s, g_a).

При кодуванні сигналу збудження кадр розбивається на 4 підкадрів по 7,5 мс. У кожному підкадрі кодуються і передаються індекси C_s (9 біт на індекс), C_a (8 біт в парних і 6 біт в непарних підкадрів на індекс), коефіцієнти підсилення g_s і $_{ga}$ (по 5 біт на кожен коефіцієнт).

У цілому кадр кодується 144 бітами, з яких 40 біт (28,8%) відводяться на кодування коефіцієнтів підсилення з використанням скалярного квантування.

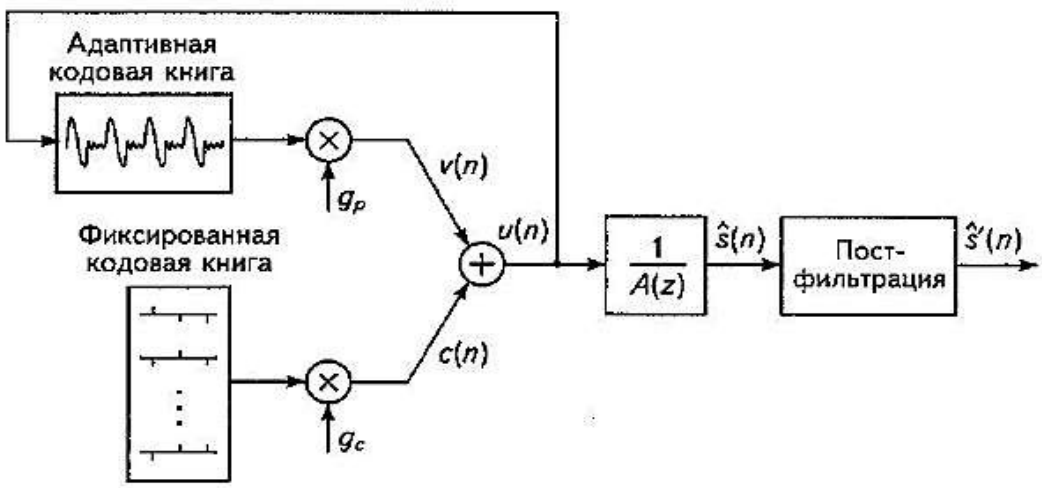


Рисунок 2.21 – CELP-кодер FS-1016

2.4.2 Побудова моделі перцептуального кодера мови. Структура кодера

В кваліфікаційній роботі пропонується наступна структурна схема CELP-кодера (рисунок 2.22), ядром якої є мультисмугова кодова книга, що складається з субсмугових кодових книг форм і коефіцієнтів підсилення.

Вхідний сигнал квантується за допомогою 16 біт / відлік з частотою дискретизації f_s = 16 кГц після чого кодується широкосмуговим кодером, структура якого представлена на рис. 2.22.

Кодер використовує блочний LP-аналіз, без інтерполяції коефіцієнтів лінійного передбачення. В якості найбільш оптимального для смуги частот 50 - 8000 Гц, був визначений p = 16 порядок фільтра-передбачення (Short-Term Predictor - STP).

Для поліпшення працездатності зважуючого фільтра (здійснення незалежного контролю форми спектральної обвідної та формантного зважування) використовується фільтр попередніх перекручень.

Кодер має восьмирівневу структуру, тобто кількість рівнів відповідає кількості оброблюваних субсмуг. Пошук здійснюється послідовно по мірі перцептуальної важливості смуг, тобто від низькочастотних до високочастотних. [5] Черговий оптимальний вектор з відповідним

коефіцієнтом посилення пропускається через послідовно з'єднані LTP-і зважуючий синтезуючий фільтр.

По закінченні циклу мінімізації помилки із залишкового зваженого оригінального сигналу віднімається внесок черговий субсмуги (рівня). Процес виконується послідовно для всіх восьми субсмуг. Після обробки всіх смуг, проводиться синтез мовного вектора по оптимальному восьмисмуговому збудливому сигналі, LTP-буфер оновлюється синтезованим вектором сигналу, а також здійснюється оновлення пам'яті зважуючого синтезуючого фільтра. Алгоритм повторюється для субфреймів, що залишилися.

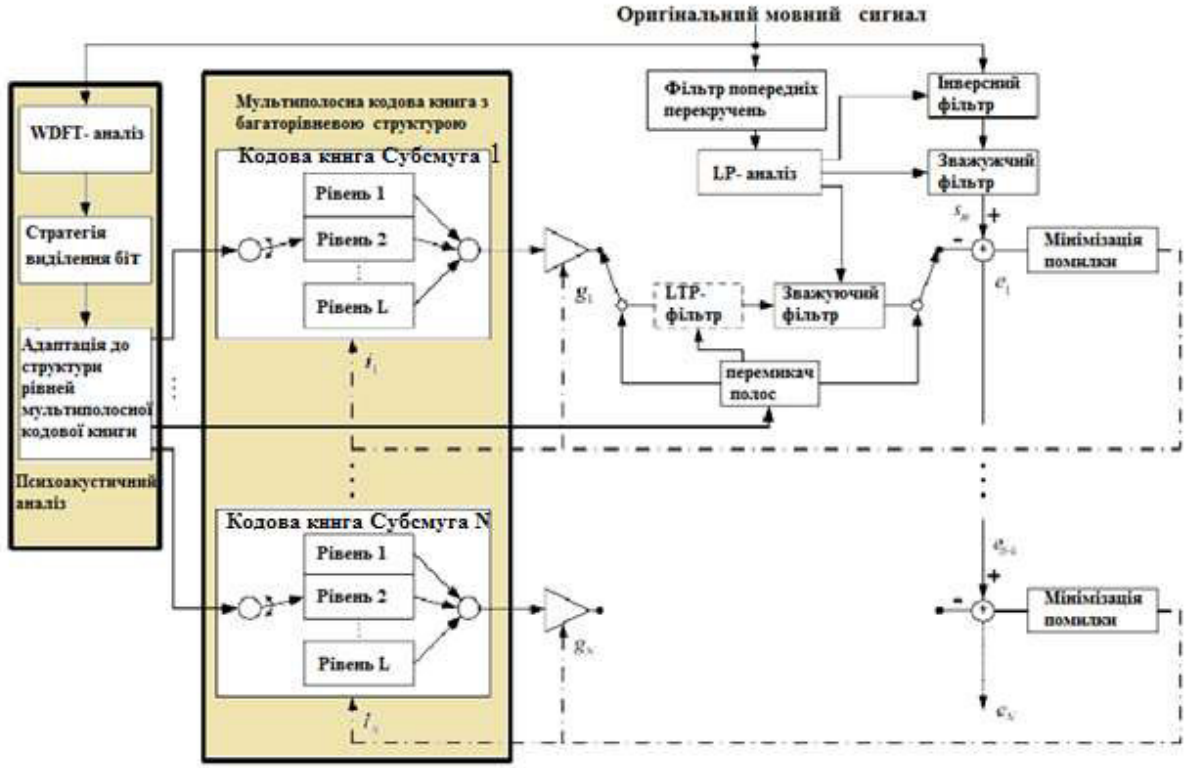


Рисунок 2.22 – Структурна схема CELP-кодера

2.4.3 Розбиття частотної осі на субсмуги

У розглянутому перцептуальному CELP-кодері з метою відображення частотно залежної поведінки аудиторної системи людини прийнята шкала барків і використовується схема розбиття на субсмуги, представлена в таблиці 2.2. Вибір субсмуг – результат інформативних тестів

прослуховування, а також оптимізації критеріїв.

Ширина смуг збільшується з частотою з метою відображення частотної вибірковості аудиторної системи людини і формантної структури мови. Для формування субсмугових навчальних множин був використаний поліфазний банк фільтрів з косинусною модуляцією і нерівномірним розбиттям на смуги, згідно таблиці 2.2 АЧХ банку фільтрів аналізу представлена на рисунку 2.23.

Необхідно відзначити, що відфільтровані кодові вектори мають кінцеву тривалість (рівну довжині субфрейма), внаслідок чого існує витік «просочування» енергії між суміжними кодовими книгами (субсмугами).

Однак, так як кодові вектори різних кодових книг близько ортогональні, послідовний пошук в кожній субсмуговій кодовій книзі забезпечує майже таку якість, що й оптимальний спільний пошук у всіх субсмугових книгах, але зі значним зниженням складності. Більше того, «просочування» спектральної енергії і-й субсмуги може бути компенсовано при квантуванні і +1-й субсмуги, тому немає необхідності в банку фільтрів з перфективною реконструкцією.

Таблиця 2.2 – Характеристика субсмуг кодера

Субсмуга	Частотний діапазон, Гц	Барки
1	100-510	4
2	510-1080	4
3	1080-1720	3
4	1720-2320	2
5	2320-3150	2
6	3150-4100	1.5
7	4100-5300	1.5
8	5300-8000	3

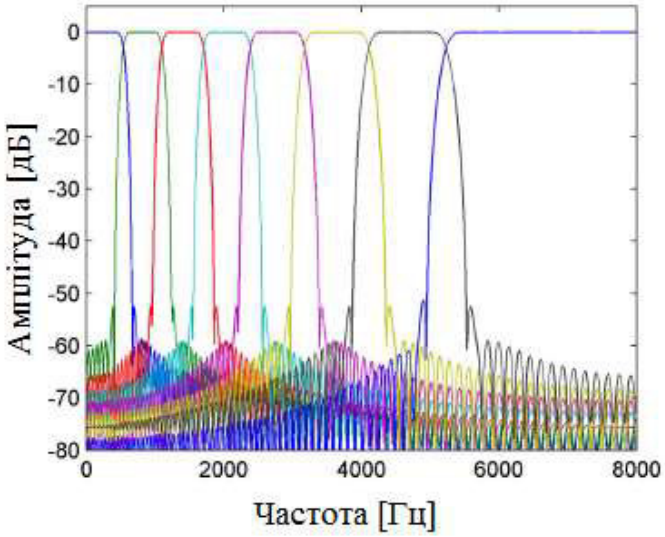


Рисунок 2.23 – АЧХ банка фільтрів

2.4.4 Реконфігурація структури кодової книги і математичне моделювання в пакеті MATLAB

Оцінка порогу маскування була підстроєна під промову з смугою 8 кГц і модифікована для використання властивостей WDFT. Основні кроки оцінки порога маскування наступні: розрахунок енергії критичних частотних смуг по WDFT-спектру потужності; згортки з функцією розповсюдження по базилярній мембрані; віднімання відносного зсуву порогу; нормалізація і порівняння з абсолютним порогом чутності [6].

Енергія розраховується в критичних частотних смугах шляхом підсумовування компонент спектру потужності в кожній критичній частотній смузі. Коефіцієнти перетворення групуються в субсмуги, в кожній з яких здійснюється оцінка порогів маскування. Як згадувалося вище, у разі WDFT ми маємо нерівномірну концентрацію енергії в околицях преутворених центральних частот.

Тому, для обчислення енергії, зосередженої в критичних частотних смугах, необхідно врахувати масштабування. За обчисленої енергії порогу маскування здійснюється визначення перцептуальної важливості кодованих субсмуг або субсмугова перцептуальна ентропія:

$$SPE_b = log_2\left(\frac{E_b}{\sqrt{\frac{6T_b}{bandwidth_b}}} + 1\right), \left[\frac{біт}{смуги}\right] \quad (2.61)$$

де E_b – енергія субсмуги; T_b – субсмуговий поріг маскування; $bandwidth_b$ – ширина субсмуги b в спектральних відліках [7].

Для того щоб перетворити перцептуальну ентропію в кожній субсмузі SPEb в глибину пошуку у відповідній субсмуговій кодовій книзі depthb, використовується масштабування вектора SPE для кожного аналізованого фрейму вхідного сигналу таким чином, щоб максимальне значення у векторі SPE_{max} відповідало максимально можливій глибині пошуку в субсмуговій кодовій книзі $depth_{bMAX}$.

Для цього обчислюється внесок кожного рівня субсмугової книги PE_per_stageb в сумарне значення субсмугової перцептуальної ентропії SPE_{max}:

$$depth_{b,MAX} \geq \max(SPE_b)$$

$$SPE_b\ PE_per_stage_b = \frac{\max(SPE_b)}{depth_{b,MAX}},$$

$$SPE_b\ PE_per_stage_b = \frac{\max(SPE_b)}{depth_{b,MAX}},$$

$$depth_b = \text{round}\,(SPE_b\ PE_per_stage_b)$$

де B – множина субсму; round – операція округлення.

Принцип зміни структури книги полягає в наступному. Оригінальний мовний сигнал (черговий кодований фрейм) надходить на вхід блоку WDFT-аналізу, в якому здійснюється оцінка субсмугових порогів маскування T_b, потім проводиться оцінка перцептуальної ентропії кожної з субсмуг кодера SPE_b. Таким чином, на виході ми маємо свого роду стратегію виділення біт або вектор SPE, що містить значення перцептуальної важливості закодованих субсмуг для поточного кадру сигналу.

Дана стратегія, згідно з алгоритмом, перетворюється у вектор depth, що містить глибину пошуку (номер рівня деталізації векторів збудження) в

субсмугових кодових книгах. При цьому якщо $SPE_b = 0$, то дана субсмуга видаляється з розгляду, тобто не кодують індекс вектора збудження і його коефіцієнт підсилення.

Вектор depth квантується по 10-бітній кодової книзі. Якщо позначити $Stage_Size_b$ [i] розмір i-го рівня субсмугової кодової книги b у векторах, то кількість біт, необхідних для кодування індексів векторів збудження в багаторівневої мультисмугової кодової книзі, складе:

$$\text{Book Indexes} = 4 * \sum^{8}_{\substack{b=1 \\ SPE_b > 0}} \log_2 (Stage_Size_b [depth_b]), [\text{біт/фрейм}], \quad (2.63)$$

а для коефіцієнтів підсилення:

$$\text{Excitation Gains} = 4 * \sum^{8}_{\substack{b=1 \\ SPE_b > 0}} 4, [\text{біт/фрейм}] \quad (2.64)$$

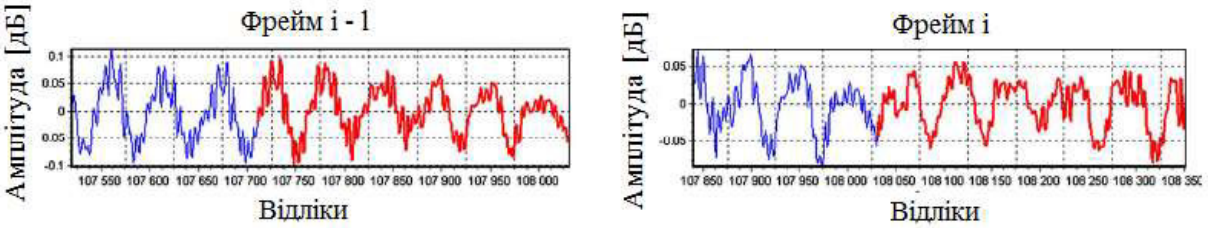
Буферизація аналізованого мовного сигналу (рис. 2.24,а) дозволяє враховувати властивості попереднього аналізованого фрейму, при цьому можна уникнути різких змін у структурі кодової книги.

На рис. 2.24,б проілюстровані пороги маскування в субсмугах для двох послідовних фреймів.

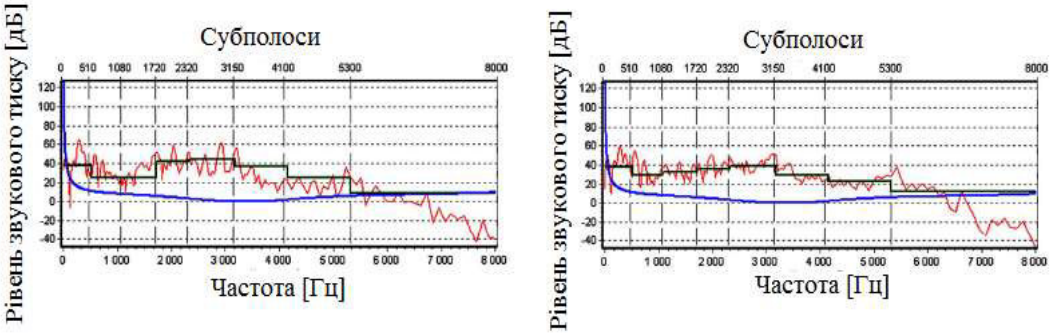
Аналіз рис. 2.24в, г, д показує, що чим менше значення міри пологості спектру субсмугового сигналу, тим більше величина відношення сигнал-маска і, отже, значення субсмугової перцептуальної ентропії.

На рис. 2.25 наведена структура мультисмугової багаторівневої кодової книги.

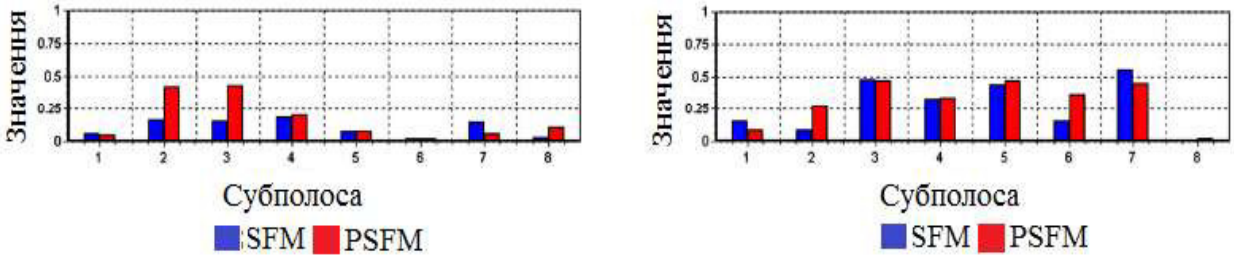
На рис. 2.26 наведено приклад зміни швидкості потоку залежно від перцептуальної важливості сигналу.



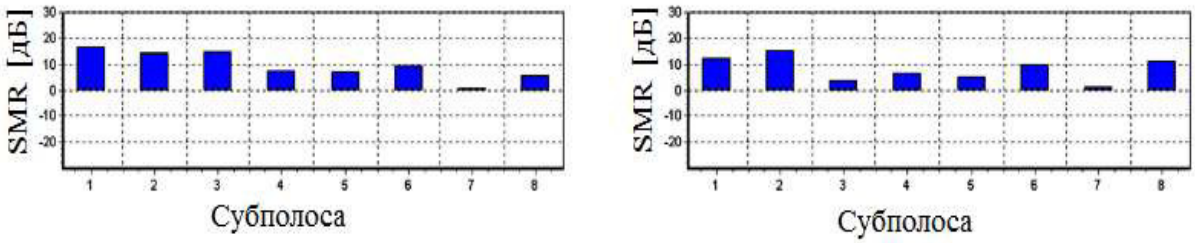
а) буфер аналізуючого мовного сигналу



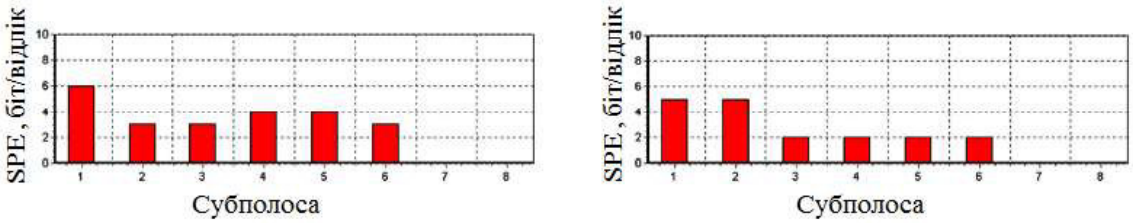
б) пороги маскування в субсмугах



в) оцінка міри пологості спектру SFM і PSFM в субсмугах



г) відношення сигнал-маска SMR в субсмугах



д) величини субсугової перцептуальної ентропії

Рисунок 2.24 – Приклад роботи психоакустичної моделі з результатами реконфігурації мультисмугової багаторівневої кодової книги

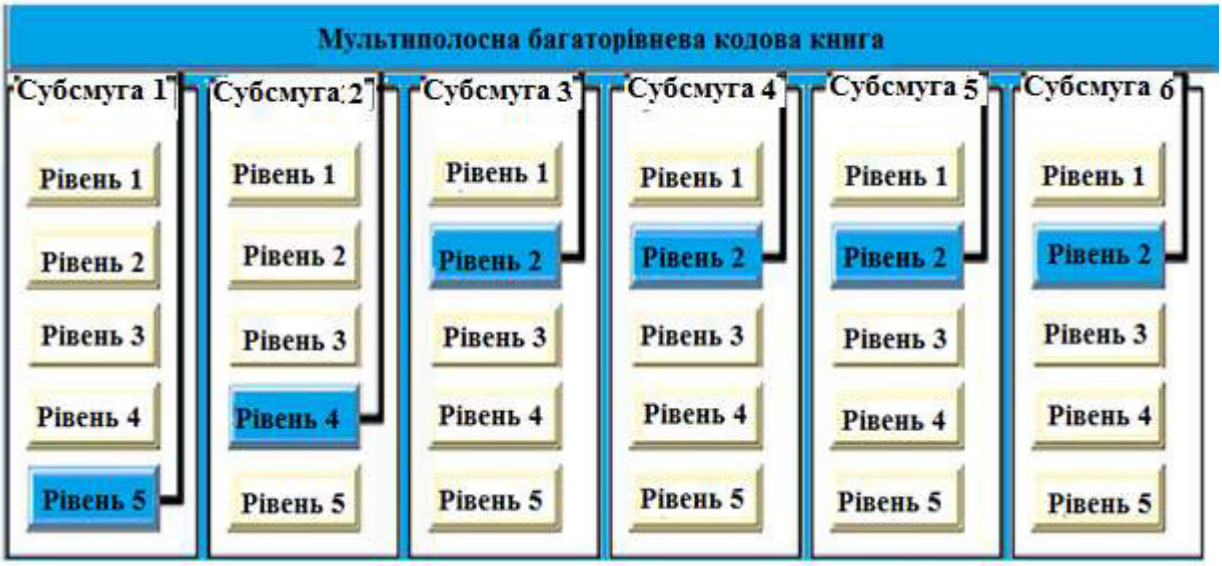


Рисунок 2.25 – Структура мультисмугової багаторівневої кодової книги

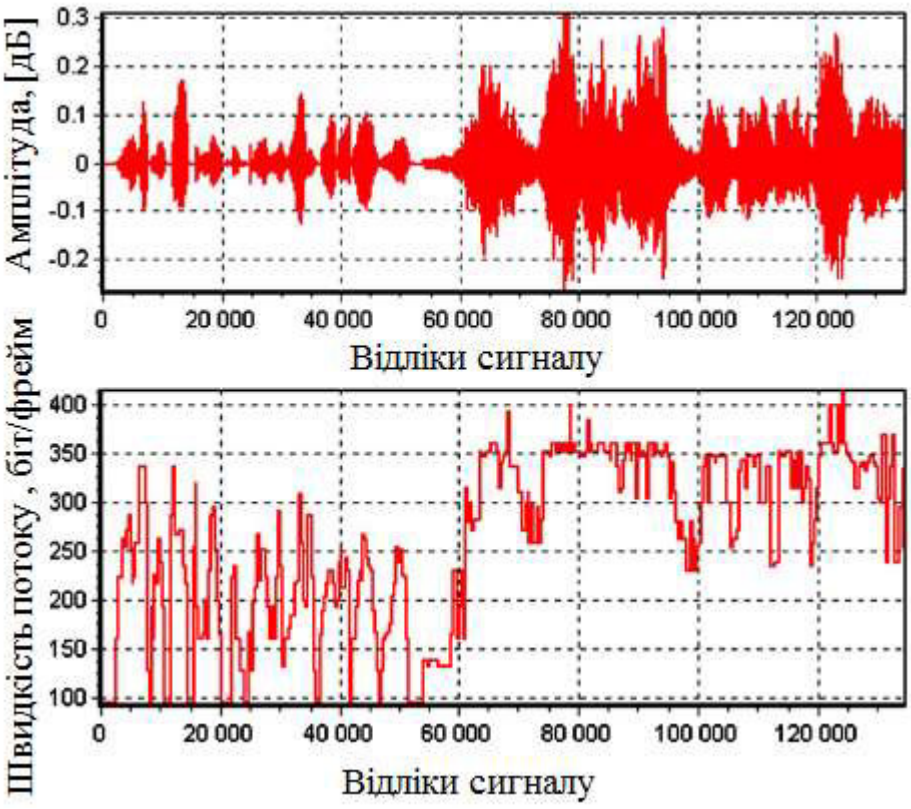


Рисунок 2.26 – Зміна швидкості потоку залежно від перцептуальної важливості сигналу: огинаюча сигналу (зверху), потік даних (знизу)

2.5 Висновок до другого розділу

За результатами дослідження можна зробити наступні висновки:

1. Завдяки вдосконаленню методу кодування лінійного передбачення з мультикодовим управлінням запропонована модель кодера дозволяє знизити об'єм переданої інформації, зберігаючи високу якість мовного сигналу.
2. Розбиття частотної осі на субсмуги забезпечує майже ту якість, що й оптимальний спільний пошук у всіх субсмугових книгах, але зі значним зниженням складності.
3. Після оптимізації обсягу кодових книг різних рівнів з урахуванням апріорних знань про перцептуальну значимість субсмуг якість закодованого сигналу підвищилася на 0,24 бала за п'ятибальною шкалою MOS.

3 ЕКОНОМІЧНИЙ РОЗДІЛ

3.1. Техніко-економічне обґрунтування створення нового алгоритм кодування мови

Необхідність кодування мовної інформації виникла не так давно, але на сьогоднішній момент, у зв'язку з бурхливим розвитком техніки зв'язку, вирішення цієї проблеми має велике значення при розробці систем зв'язку.

Підвищення якості веде до збільшення обсягу інформації, для передачі якої необхідна велика ширина каналу передачі. У зв'язку з цим виникає необхідність розробки і створення методів кодування з високим ступенем стиснення при збереженні високої якості.

На сьогоднішній день в кодуванні мови існує проблема великої обчислювальної складності алгоритмів, а також не використаний весь потенціал методів кодування для отримання більш якісного звучання. Тому метою дипломної роботи є підвищення якості мовного сигналу без збільшення швидкості передачі за рахунок реконфігурації мультисмугової кодової книги.

3.2 Визначення трудомісткості розробки алгоритму

Трудомісткість розробки покращеного алгоритму кодування може бути розрахована на основі системи моделей з різною точністю оцінки.

Трудомісткість розробки програмного забезпечення можна розрахувати за формулою:

$$t = t_{тз} + t_{И} + t_{А} + t_{П} + t_{Д}, \text{ люд/год} \qquad (3.1)$$

де, $t_{тз}$ – витрати праці на підготовку і опис технічного завдання (50 люд/год);

$t_{И}$ – витрати праці на дослідження методів кодування;

$t_{А}$ – витрати праці на розробку блок-схеми алгоритму;

$t_П$ – витрати праці на створення нової структурної схеми кодера;

$t_Д$ – витрати праці на підготовку документації.

Складові витрати праці визначаються через умовну кількість елементарних операцій в алгоритмі.

Умовне число елементарних операцій в алгоритмі:

$$Q = q \cdot C \cdot (1 + p), \tag{3.2}$$

де , q – число базових елементарних операцій алгоритму (1500);

c – коефіцієнт складності розроблюваного алгоритму (1.7);

p – коефіцієнт корекції алгоритму в ході його розробки (0.15).

Звідси умовне число елементарних операцій алгоритму:

$$Q = 1{,}7 \cdot 1500 \cdot (1 + 0{,}15) = 2932{,}5, \tag{3.3}$$

Витрати праці на вивчення опису завдання $t_И$ визначаються з урахування уточнення опису і кваліфікації програміста:

$$t_и = \frac{Q \cdot B}{80 \cdot K}, \tag{3.4}$$

де, B – коефіцієнт збільшення витрат праці внаслідок недостатнього опису завдання;

K – коефіцієнт кваліфікації розробника, що визначається залежно від стажу роботи з даної спеціальності. При стажі роботи від 3 до 5 років, він складає 1.

Приймемо збільшення витрат праці внаслідок недостатнього опису завдання не більше 50% (B = 1.5). З урахуванням коефіцієнта кваліфікації K = 1 отримуємо витрати праці на вивчення опису завдання:

$$t_И = (2935{,}5 \cdot 1{,}5) / (80 \cdot 1) = 54{,}98 \text{ люд/год.} \tag{3.5}$$

Витрати праці на разробку покращеного алгоритму визначаються по формулі:

$$t_А = \frac{Q}{20K}, \tag{3.6}$$

де Q – умовне число елементарних операцій в алгоритмі; K – коефіцієнт кваліфікації розробника.

Підставивши відповідні значення в формулу (3.2), отримаємо:

$$t_A = 2932,5/ (20 \cdot 1) = 146,25 \text{ люд/год} \quad (3.7)$$

Витрати на створення нової структурної схеми кодера

$$t_{\ddot{I}} = \frac{Q}{25K}, \quad (3.8)$$

$$t_П = 2932,5 / (25 \cdot 1) = 117,3 \text{ люд/год.}$$

Витрати на підготовку документації визначаються по формулі

$$t_Д = t_{др} + t_{до}, \quad (3.9)$$

де, $t_{др}$ – трудомісткість підготовки матеріалів і рукопису;

$t_{до}$ – трудомісткість редагування, друку і оформленння документації.

$$t_{др} = \frac{Q}{20 \cdot K}, \quad (3.10)$$

$$t_{до} = 0,75 \cdot t_{др}. \quad (3.11)$$

Підставивши відповідні значения, отримаємо:

$$t_{др} = 2932,5 / (20 \cdot 1) = 146,62 \text{ люд/год.}$$

$$t_{до} = 0,75 \cdot 146,62 = 109,96 \text{ люд/год.}$$

$$t_Д = 146,62 + 109,96 = 256,58 \text{ люд/год.}$$

Повертаючись до формули (3.1), отримуємо повну оцінку трудомісткості розробки алгоритму:

$$t = 40 + 54,98 + 146,25 + 117,3 + 256,58 = 615,11 \text{ люд/год.}$$

3.3. Витрати на створення алгоритму кодування

Витрати на створення алгоритму ($К_{по}$) включають витрати на зарабітну плату розробку ($З_{зп}$) і вартість машинного часу, необхідного для його розробки на ЕОМ ($З_{мв}$):

$$К_{по} = З_{зп} + З_{мв}, \text{ грн.} \quad (3.10)$$

Зарабітна плата розробника визначається по формулі:

$$З_{зп} = t \cdot С_{пр}, \text{ грн} \quad (3.11)$$

де

t – загальна трудомісткість розробки програми, люд/год;

$С_{пр}$ - середня погодинна зарабітна плата розробника з нарахуваннями.

З урахуванням того, що середня почасова зарплата розробника становить 25 грн/год, отримуємо:

$$З_{зп} = 615,11 \cdot 25 = 15377,75 \text{ грн.}$$

Вартість машинного часу, необходного для створення алгоритму на ЕОМ, визначається по формулі

$$З_{зп} = t_{рс} \cdot С_{мч}, \text{ грн} \quad (3.12)$$

де

$t_{рс}$ – трудомісткість розробки алгоритму створення схеми на ЕОМ, год;

$С_{мч}$ – вартість машинного часу ЕОМ (4 грн/час).

Підставивши в формулу (3.12) відповідні значення, визначимо вартість машинного часу, необходного для створення алгоритму:

$$З_{мв} = 586,5 \cdot 4 = 2346 \text{ грн.}$$

Звідси затрати на розробку покращенного кодуючого методу:

$$К_{по} = 15377,75 + 2346 = 17723,75 \text{ грн.}$$

Очікуваний період розробки алгоритму і кодируючої схеми:

$$T = \frac{t}{B_k \cdot F_p}, \text{ міс,} \quad (3.12)$$

де

B_k – число розробників (рівно 1),

$F_р$ – місячний фонд рабочого часу (при 40-годинному робочому тижні $F_р$=176 годин).

Підставивши відповідні значення, отримаємо:

$$Т = 615,11 / 1 \cdot 176 \approx 3,4 \text{ міс.}$$

3.4 Визначення економічної переваги розробленого алгоритму

Провівши аналіз витрат на розробку нового способу кодування та існуючих, було виявлено що існує кілька аналогів, вартість яких в середньому складає від 6000 грн до 7200 грн.

В результаті розрахунку витрат на розробку даного алгоритму, пропонується призначити вартість - 5800 грн., що на 1400 грн. дешевше його аналогів.

Провайдери зацікавлені в менш ресурсномістких алгоритмах кодування. Ця обставина дає перевагу розробленому алгоритму і гарантує високий попит на його використання.

3.5 Висновок до третього розділу

Висновок, який можна зробити в даному розділі полягає в наступному: розроблений алгоритм кодування дозволяє отримати більш якісний результат в порівнянні з попередніми; апаратна або програмна реалізація цього алгоритму буде дешевша на 1400 грн., так як знижена обчислювальна складність методу кодування.

ВИСНОВКИ

У кваліфікаційній роботі описані ефективні структури і методи проектування банків цифрових фільтрів.

Докладно розглянуті ДПФ-модульовані банки фільтрів. Запропоновано застосовувати для обробки мови дискретне перетворення Фур’є з нерівномірним частотним дозволом. Це дозволяє апроксимувати психоакустичні шкали для побудови перцептуально мотивованих систем стиснення і передачі мовної інформації, а також систем підвищення якості мови.

Завдяки вдосконаленню методу кодування лінійного передбачення з мультикодовим управлінням запропонована модель кодера, яка дозволяє знизити об'єм переданої інформації, зберігаючи високу якість мовного сигналу.

Розбиття частотної осі на субсмуги забезпечує майже ту якість, що й оптимальний спільний пошук у всіх субсмугових книгах, але зі значним зниженням складності.

Після оптимізації об’єму кодових книг різних рівнів з урахуванням апріорних знань про перцептуальну значимість субсмуг якість закодованого сигналу підвищилася на 0,24 бала за п'ятибальною шкалою MOS.

Виконана перевірка запропонованого способу мовного кодування в телекомунікаційній системі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Электросвязь [Электронный ресурс]. – Режим доступу: http://ru.wikipedia.org/wiki/Электросвязь

2. Telecommunication [Электронный ресурс]. – Режим доступу: http://edu.dvgups.ru/METDOC/GDTRAN/YAT/TELECOMM/UPIOI/METOD/MU_RGR/frame/1.htm

3. CELP [Электронный ресурс]. – Режим доступу: http://ru.wikipedia.org/wiki/CELP

4. Federal Standard 1016. Telecommunications: Analog to Digital Conversion of Radio Voice by 4,800 bit/second Code Excited Linear Prediction (CELP). February 14, 1991.

5. Основы психоакустики [Электронный ресурс]. – Режим доступу: http://nsk.jagannath.ru/users_files/books/Osnovy_psihoakustiki.

6. Кодирование сигналов [Электронный ресурс]. – Режим доступу: http://kit-e.ru/articles/circuit/2010_07_150.php

7. Способы обработки и кодирования сигналов [Электронный ресурс]. – Режим доступу: http://acustika.info/category/sposobi-i-ustroystva-obrabotki-i-kodirovaniya-zvukovih-signalov/page/11

8. Компрессия звука [Электронный ресурс]. – Режим доступу: http://www.compression-pointers.ru

9. Сапожков М.А., Михайлов В.Г. Вокодерная связь. –М.: Радио и связь, 1983.

10. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. М.: Радио и связь, 2000.

11. Кузьмин И.В. Основы теории информации и кодирования. – Минск: Вышэйш. шк., 1986.

12. Хемминг Р.В. Теория информации и теория кодирования. - М.: Радио и связь, 1983.

13. Federal Standard 1016. Telecommunications: Analog to Digital Conversion of Radio Voice by 4,800 bit/second Code Excited Linear Prediction (CELP). February 14, 1991.

14. Шульгин В.И. Основы теории передачи информации. – Харьков «ХАИ» , 2003.

15. ISO/IEC JTC1/SC29/WG11 NO803, MPEG, International Standard IS 13818-3 Information Technology – Generic Coding of Moving Pictures and Associated Audio: Audio, 11th November 1994.

ДОДАТОК А. Перелік документів КВАЛІФІКАЦІЙНОЇ РОБОТИ

1 Пояснювальна записка на 87 сторінках.

2 Матеріали кваліфікаційної роботи на оптичному носії:

- Пояснювальна записка Ковальська О.О.doc
- Презентація.pptx

ДОДАТОК Б. Відгук керівника економічного розділу

Керівник розділу _____ Романюк Н.М.

(підпис) (прізвище, ініціали)

ДОДАТОК В. ВІДГУК КЕРІВНИКА КВАЛІФІКАЦІЙНОЇ РОБОТИ

ВІДГУК

на кваліфікаційну роботу ступеня бакалавр студентки групи 172-16зск-1
Ковальської О.О. на тему:
«Удосконалення способу мовного кодування в телекомунікаційній системі»

Мета кваліфікаційної роботи – підвищення якості мовного сигналу без збільшення швидкості передачі за рахунок реконфігурації мультисмугової кодової книги.

Сучасний етап розвитку суспільства характеризується збільшенням потоків мовних повідомлень, що передаються по телекомунікаційним системам та мережам, що мають обмежену пропускну спроможність. Великі труднощі виникають також при вирішенні задачі зберігання мовних даних, зростаючий потік яких потребує все більших обсягів пам'яті запам'ятовуючих пристроїв.

У різних країнах світу, для вирішення цієї проблеми інтенсивно розвивається напрям, що полягає в стисненні обсягів мовних даних. Цей напрямок особливо привабливо з економічної точки зору, так як не вимагає створення нових дорогих виробничих потужностей, але призначено для ефективного використання наявних ресурсів систем і мереж.

Один з перспективних напрямів є удосконалення способу мовного кодування в телекомунікаційній системі на основі широкосмугових CELP кодерів мови.

Оформлення кваліфікаційної роботи виконано на відповідному рівні і відповідає вимогам, що пред'являються до робіт даної кваліфікації. Рівень запозичень у кваліфікаційній роботі відповідає вимогам «Положення про систему виявлення та запобігання плагіату». В цілому кваліфікаційна робота ступеня бакалавра повністю задовольняє вимогам, що пред'являються і заслуговує оцінки «відмінно», а її автор, Ковальська О.О. присвоєння кваліфікації технічного фахівця в галузі електроніки та телекомунікацій.

Керівник кваліфікаційної роботи, доц. Магро В.І.

ДОДАТОК Г. Рецензія

РЕЦЕНЗІЯ

на кваліфікаційну роботу ступеня бакалавр студентки групи 172-16зск-1
Ковальської О.О. на тему:
«Удосконалення способу мовного кодування в телекомунікаційній системі»

Представлена на рецензію кваліфікаційна робота виконана на 86 машинописних сторінках і складається з вступу, трьох розділів і висновку. Кваліфікаційна робота студентки групи 172-16зск-1 Ковальської О.О. виконана в повному обсязі відповідно до завдання.

У кваліфікаційній роботі описані ефективні структури і методи проектування банків цифрових фільтрів. Виконана реалізація фільтрів в середовищі MATLAB.

В кваліфікаційній роботі запропоновано технічне рішення, що відноситься до засобів формування звукового сигналу з поліпшеною для сприйняття якістю. Технічним результатом є підвищення якості звуковідтворення. Зважаючи на це тема кваліфікаційної роботи є актуальною.

Позитивні сторони:

1. Зміст кваліфікаційної роботи відповідає завданню. Робота, яка виконана Ковальською Оленою Олександрівною, показала високий рівень знань і ступінь підготовленості її до майбутньої роботи з фаху.
2. Підготовлений в кваліфікаційній роботі огляд методів використання цифрових банків фільтрів виконаний в повному обсязі.
3. Текст викладений грамотно, ясно, послідовно. Графічний матеріал оформлений якісно. Широко використовується науково-технічна література.

Недоліки:

1. Огляд методів кодування виконаний в трохи завищеному об’ємі.
2. Слід було описати алгоритм запропонованого способу обробки сигналу.

В цілому кваліфікаційна робота ступеня бакалавр повністю задовольняє вимогам, що пред'являються і заслуговує оцінки «відмінно », а її автор, Ковальська Олена Олександрівна, присвоєння кваліфікації технічного фахівця в галузі електроніки та телекомунікацій.

Рецензент,

_____
(підпис)