

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

Інститут електроенергетики
Факультет інформаційних технологій
Кафедра безпеки інформації та телекомунікацій

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня бакалавра

студента Синій Дмитро Сергійович

академічної групи 172-18ск-1

спеціальності 172 Телекомунікації та радіотехніка

спеціалізації¹

за освітньо-професійною програмою Телекомунікації та радіотехніка

на тему Векторне квантування параметрів лінійного
передбачення мовного сигналу

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

Дніпро
2021

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

« _____ » _____ 20__ року

ЗАВДАННЯ
на кваліфікаційну роботу
ступеня бакалавра

студенту _____ *Синій Дмитро Сергійович* _____ академічної групи *172-18ск-1*
(прізвище ім'я по-батькові) (шифр)

спеціальності *172 Телекомунікації та радіотехніка*

за освітньо-професійною програмою *Телекомунікації та радіотехніка*

на тему _____ *Векторне квантування параметрів лінійного*
передбачення мовного сигналу

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

Розділ	Зміст	Термін виконання
Розділ 1	Аналіз основних положень формування і обробки мовних сигналів, а також існуючих підходів до векторного квантування параметрів лінійного передбачення.	25.02.2021 – 31.03.2021
Розділ 2	Розробка підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови та оцінка його ефективності.	01.04.2021 – 12.05.2021
Розділ 3	Розрахунки витрат на розробку підходу до векторного квантування параметрів лінійного передбачення.	13.05.2021 – 09.06.2021

Завдання видано _____

(підпис керівника)

Герасіна О.В.

(прізвище, ініціали)

Дата видачі: _____

Дата подання до екзаменаційної комісії: _____

Прийнято до виконання _____

(підпис студента)

Синій Д.С.

(прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 72 с., 20 рис., 1 таблиця, 4 додатки, 25 джерел.

Об'єкт розробки – кадри мовного сигналу.

Предмет розробки – підхід до векторного квантування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу.

Мета кваліфікаційної роботи – зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування.

Наукова новизна результатів полягає у застосуванні схеми передбачення із адаптацією до основного тону мови при класифікації кадру мовного сигналу як стаціонарного локалізованого.

У першому розділі проаналізовано основні положення формування і обробки мовних сигналів, а також існуючі підходи до векторного квантування параметрів лінійного передбачення.

У спеціальній частині роботи запропоновано підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови та оцінено його ефективність. За наслідками досліджень зроблено висновки щодо рішення поставленої задачі.

У економічному розділі виконані розрахунки витрат на розробку підходу до векторного квантування параметрів лінійного передбачення.

СПЕКТРАЛЬНЕ СПОТВОРЕННЯ, ВЕКТОРНЕ КВАНТУВАННЯ,
СТАЦІОНАРНИЙ ВОКАЛІЗОВАНИЙ КАДР МОВНОГО СИГНАЛУ,
КОДУВАННЯ З ЛІНІЙНИМ ПЕРЕДБАЧЕННЯМ, АДАПТАЦІЯ ДО
ОСНОВНОГО ТОНУ МОВИ, ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ

РЕФЕРАТ

Пояснительная записка 72 с., 20 рис., 1 таблица, 4 приложения, 25 источника.

Объект разработки – кадры речевого сигнала.

Предмет разработки – подход к векторному квантованию параметров линейного предсказания стационарных вокализованных кадров речевого сигнала.

Цель квалификационной работы – снижение количества бит, выделяемых для кодирования параметров линейного предсказания стационарных вокализованных кадров речевого сигнала при минимальной погрешности квантования.

Научная новизна заключается в применении схемы предсказания с адаптацией к основному тону речи при классификации кадра речевого сигнала как стационарного локализованного.

В первой главе проанализированы основные положения формирования и обработки речевых сигналов, а также существующие подходы к векторному квантованию параметров линейного предсказания.

В специальной части работы предложен подход к векторному квантованию параметров линейного предсказания с адаптацией к основному тону языка и оценена его эффективность. По результатам исследований сделаны выводы относительно решения поставленной задачи.

В экономическом разделе выполнены расчеты затрат на разработку подхода к векторному квантованию параметров линейного предсказания.

СПЕКТРАЛЬНОЕ ИСКАЖЕНИЕ, ВЕКТОРНОЕ КВАНТОВАНИЕ, СТАЦИОНАРНЫЙ ВОКАЛИЗОВАННЫЙ КАДР РЕЧЕВОГО СИГНАЛА, КОДИРОВАНИЕ С ЛИНЕЙНЫМ ПРЕДСКАЗАНИЕМ, АДАПТАЦИЯ К ОСНОВНОМУ ТОНУ ЯЗЫКА, ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

ABSTRACT

Explanatory note: p. 72, fig. 20, tab. 1, 4 additions, 25 sources.

The object of development is frames of a speech signal.

The subject of development is an approach to vector quantization of linear prediction parameters of stationary voiced frames of a speech signal.

The purpose of the qualification work is to reduce the number of bits allocated for encoding linear prediction parameters of stationary voiced frames of a speech signal with a minimum quantization error.

The scientific novelty lies in the application of a prediction scheme with adaptation to the fundamental tone of speech when classifying a frame of a speech signal as stationary localized.

The first chapter analyzes the basic provisions of the formation and processing of speech signals, as well as existing approaches to vector quantization of linear prediction parameters.

In a special part of the work, an approach to vector quantization of linear prediction parameters with adaptation to the main tone of the language is proposed and its efficiency is estimated. Based on the results of the research, conclusions were drawn regarding the solution of the task.

The economic section calculates the costs of developing an approach to vector quantization of linear prediction parameters.

SPECTRAL DISTORTION, VECTOR QUANTIZATION, STATIONARY VOCALIZED FRAME OF SPEECH SIGNAL, CODING WITH LINEAR PREDICTIONS, ADAPTATION TO THE BASIC TONE OF THE LANGUAGE, SIMULATION MODELING

СПИСОК УМОВНИХ СКОРОЧЕНЬ

- АМ – Амплітудний модулятор;
- ВК – Векторне квантування;
- ГМС – Генератор мовного сигналу;
- ІКМ – Імпульсно-кодова модуляція;
- КК – Кодова комбінація;
- КЛП – Коефіцієнт лінійного передбачення;
- ЛСЧ – Лінійна спектральна частота;
- ЛП – Лінійне передбачення;
- ОТ – Основний тон;
- СФ – Смуговий фільтр;
- ФНЧ – Фільтр нижніх частот;
- LPC – Linear Prediction Coding – Кодування з лінійним передбаченням;
- RELП – Residual Excited Linear Prediction – Лінійне передбачення із залишковим збудженням;
- VELP – Voice Excited Linear Prediction – Лінійне передбачення із голосовим збудженням;
- VEV – Voice Excited Vocoder – Вокодер із голосовим збудженням.

ЗМІСТ

	с.
ВСТУП.....	9
1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ.....	11
1.1 Характеристики безперервних джерел та одержувачів повідомлень	11
1.2 Компресія мовних сигналів	17
1.2.1. Елементи теорії інформації в застосуванні до мовного сигналу	17
1.2.2 Методи безпосередньої компресії	21
1.3 Параметрична компресія мовних сигналів.....	30
1.3.1 Загальні положення.....	30
1.3.2 Смугові вокодери	33
1.4 Існуючі підходи до векторного квантування параметрів лінійного передбачення	37
1.5 Висновок. Постановка задачі.....	46
2 СПЕЦІАЛЬНА ЧАСТИНА.....	48
2.1 Підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови	48
2.2 Оцінка ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови.....	55
2.3 Висновок	58
3 ЕКОНОМІЧНИЙ РОЗДІЛ.....	60
3.1 Визначення трудомісткості векторного квантування параметрів лінійного передбачення	60
3.2 Визначення середньої заробітної плати спеціаліста в галузі телекомунікацій.....	61
3.3 Розрахунок витрат на заробітну плату при дослідженні векторного квантування параметрів лінійного передбачення.....	62
3.4 Розрахунок витрат на необхідне програмне та апаратне забезпечення	62
3.5 Висновок	63

	8
ВИСНОВКИ.....	64
ПЕРЕЛІК ПОСИЛАНЬ	66
ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи	69
ДОДАТОК Б. Перелік документів на оптичному носії.....	70
ДОДАТОК В. Відгук керівника економічного розділу.....	71
ДОДАТОК Г. Відгук керівника кваліфікаційної роботи	72

ВСТУП

Бурхливий розвиток засобів зв'язку в останній час призвів до корінної зміни як елементів бази так і принципів передавання, прийому та обробки інформації. Особливо це стосується проблем передавання мовних повідомлень. Наразі розробляються все нові й нові високоефективні алгоритми цифрової обробки мови різного призначення. Найбільш бурхливо розвивається галузь цифрового передавання мови, яка являє собою мобільний радіозв'язок, Інтернет-телефонію, у тому числі із застосуванням радіоканалів, та інші системи [1].

Наразі у зв'язку з активним розвитком техніки зв'язку, особливо мобільного, рішення задачі кодування мовної інформації має велике значення.

При кодуванні з лінійним передбаченням (Linear Prediction Coding – LPC) моделюються різні параметри мови, які передаються замість відліків або їх різниці, що вимагають значно більшої пропускної здатності каналу. LPC працює з блоками відліків. Для кожного блоку алгоритм LPC обчислює і передає частоту основного тону, його амплітуду, «прапор» мовного або немовного походження сигналу тощо.

При такому підході до кодування мови, по-перше, зростають вимоги до обчислювальних потужностей мікропроцесорів, які використовуються для обробки сигналу, а по-друге, збільшується затримка при передачі, оскільки кодування застосовується не до окремих значень, а до деякого їх набору, який перед початком перетворення слід накопичити у певному буфері. Слід підкреслити, що затримка в передачі мови при використанні LPC методу пов'язана не тільки з необхідністю обробки цифрового сигналу (цю затримку можна зменшувати, збільшуючи потужність процесора), але й безпосередньо впливає зі специфіки методу.

LPC метод дозволяє досягати дуже високих ступенів стиснення, яким відповідає смуга пропускання 2,4 або 4,8 кбіт/с, однак, при цьому якість звуку виходить низькою.

Отже, існує проблема недостатньої продуктивності комп'ютерних систем, призначених для обробки мультимедійної, зокрема голосової інформації. Зменшення обсягів даних, необхідних для коректного відтворення цифрового мовного сигналу, дозволяє істотно підвищити ефективність їх роботи. Наразі ступінь стиснення забезпечують параметричні методи з подальшим квантуванням отриманих параметрів. На етапі параметризації сигнал розбивають на кадри однакової довжини, як правило, по 10-30 мс, і для кожного з них обчислюють деякий вектор параметрів. У сучасних системах стиснення мовних сигналів стандартом де-факто стало обчислення параметрів на основі моделі лінійного передбачення десятого порядку. На етапі квантування отриманий вектор замінюють найближчим до вхідного представником репрезентативною сукупністю параметрів, що зберігаються в кодових книгах.

Таким чином, вдосконалення підходів до векторного квантування параметрів лінійного передбачення наразі є актуальною задачею.

Метою роботи є зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування.

Постановка задачі:

- проаналізувати основні положення формування і обробки мовних сигналів;
- провести аналіз існуючих підходів до векторного квантування параметрів лінійного передбачення;
- запропонувати підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови;
- оцінити ефективність розробленого підходу.

1 СТАН ПИТАННЯ. ПОСТАНОВКА ЗАДАЧІ

1.1 Характеристики безперервних джерел та одержувачів повідомлень

Джерелом безперервних повідомлень є голосовий апарат людини, а мова вважається самим повідомленням. Одержувачем мовного повідомлення є слуховий апарат людини. Системи зв'язку, в яких використовуються мовні повідомлення, ще називають телефонними системами [1-12].

Характерною особливістю систем передачі мовних повідомлень вважають наявність у них акустичних елементів (голосовий та слуховий апарати людини, мікрофон та телефон).

Складовими елементами мовних повідомлень є слова та фрази. Слова, в свою чергу, складаються з більш простих елементів мови – складів та звуків, характерних для тієї чи іншої мови. Точного визначення терміна «звук мови» не існує. Його можна порівняти з великою літерою. Залежно від мови, місця наголосу кожний звук може мати різні відтінки. Так, нараховується декілька тисяч звуків, які відрізняються за суб'єктивним сприйняттям один від одного. Проте, все ж вдається класифікувати всі звуки та їх варіанти. А якщо ще й виділити типові ознаки звуків, то їх кількість вдається скоротити до 40...42. Такі типізовані звуки мови носять назву фонем. Кожна фонема має свої характерні ознаки, які легко розрізнити на слух. Але під час навіть найточнішого наголосу її в послідовній мові внаслідок впливу сусідніх звуків вона може набувати тих чи інших відтінків. Це будуть варіанти фонем за місцем у слові або так звані фоноїди [1].

Структура звуку досить складна та характеризується цілою низкою складових звукових частот. При цьому розрізняють голосні та приголосні звуки.

Спектр голосних звуків має дискретний характер і складається з основного тону та його гармонік. Математичне очікування частоти основного тону F_{OT} для чоловічого голосу складає 120 Гц, а для жіночого – 240 Гц. Під час

промови повідомлень різними особами відбувається зміна частоти основного тону $\left(\frac{\partial F_{OT}}{\partial t}\right)$ навколо математичного очікування, що й визначає особу, яка веде розмову. Таким чином, кожна людина розмовляє з характерною для неї частотою основного тону F_{OT} та похідною $\frac{\partial F_{OT}}{\partial t}$.

Для глухих приголосних та мови пошепки характерний суцільний спектр звуків. Частотні спектри дзвінких приголосних звуків мають комбінований дискретний та суцільний вигляди.

Типова осцилограма голосного звуку показана на рис. 1.1.

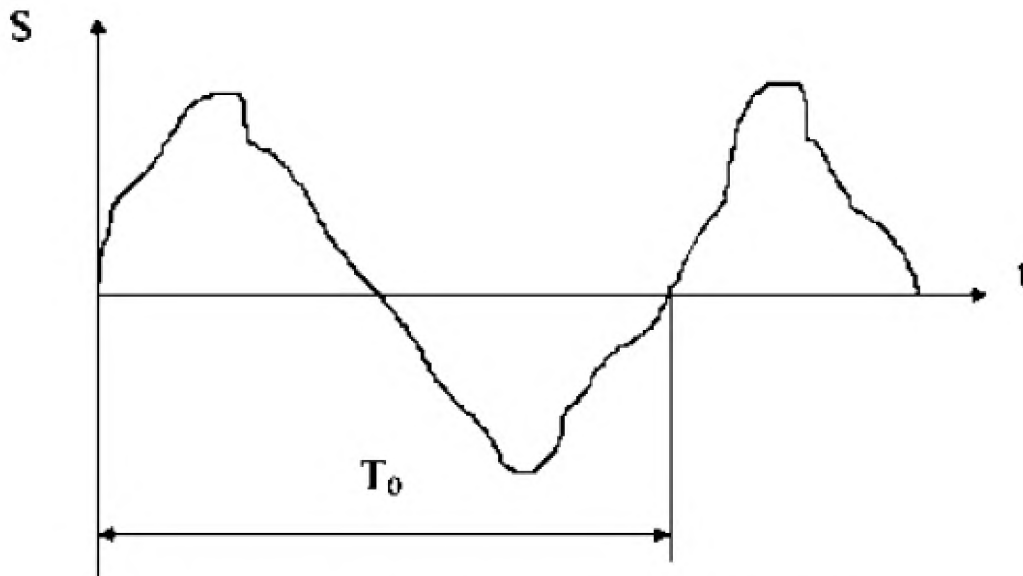


Рисунок 1.1 – Осцилограма голосного звуку

В спектрі звуків утворюються області концентрації енергії, які носять назву формант. Іншими словами, формантою називають групу суміжних частотних складових спектра звуку, рівні яких більші, ніж у сусідніх складових. Для голосних звуків характерно до трьох яскраво виражених формант, які знаходяться в діапазонах: 300...1000 Гц (це перша форманта), 1000...2000 Гц (друга форманта) та 2000...3500 Гц (третья форманта) (рис. 1.2). Глухі приголосні мають лише одну форманту з досить розмитим екстремумом у діапазоні частот понад 1200 Гц. Слід зазначити, що кожній фонемі відповідає

свій формантний рисунок. Це означає, що мозок людини аналізує розташування формант та їхню інтенсивність, отримує інформацію з мовного повідомлення.

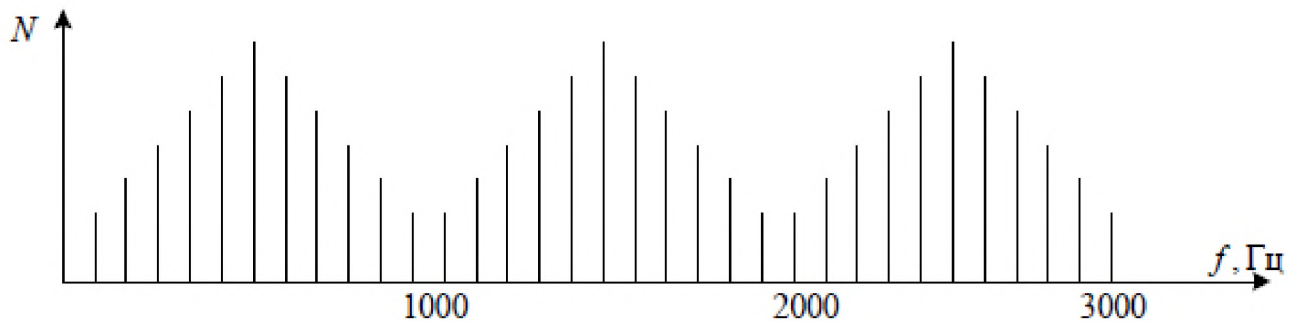


Рисунок 1.2 – Спектр голосного звуку

Так звані вибухові звуки (п, к, т) не мають особистого формантного рисунку, а являють собою ту чи іншу зміну попередніх та наступних звуків.

Таким чином, діапазон частот, в якому розташовані форманти, які несуть інформацію про звуки, що промовляються, досить вузький. Він складає від 300 до 3500 Гц і носить назву телефонного каналу. Хоч під час розмови утворюються звуки в смузі частот від 100 до 9000 Гц, з точки зору отримання інформації немає ніякого сенсу робити телефонний канал ширше від 300...3500 Гц.

Все це стосується професійних систем зв'язку. В системах же радіомовлення та телебачення намагаються передати всі частоти, які утворюються під час розмови, співу, музичного супроводження тощо.

У професійному зв'язку завжди визначають темп мови, він складає близько 10 фонем/с. У разі такого темпу (швидкості проголошення фонем) інтервал існування звуків складає 20...300 мс, при цьому у голосних він більший, ніж у приголосних. Найкоротші звуки – це вибухові, вони звучать 20...30 мс.

Під час розмови утворюється звукова хвиля, яка зумовлює відповідний звуковий тиск. Його позначають літерою P та вимірюють у паскалях (Па). Відчуття гучності пов'язано з силою звуку, яку позначають літерою I та

вимірюють у Вт/м^2 . Сила звуку 2 м Вт еквівалентна потужності, а звуковий тиск – напрузі електричних коливань, тому можна записати, що $I \equiv P^2$.

Рівень інтенсивності звуків мови (N) вимірюють у децибелах (дБ). При цьому за одиницю, з якою порівнюють звукові коливання, приймають мінімальну силу I_0 звукових коливань, що відчуває вухо людини з середнім слухом. Ця сила на середніх частотах (600...800) Гц дорівнює $I_0=10^{-2}$ Вт/м^2 або $P_0=2 \cdot 10^{-5}$ Па [1].

Тоді рівень будь-якого звукового коливання силою I та звуковим тиском P буде дорівнювати:

$$N = 10 \lg \frac{I}{I_0} = 20 \lg \frac{P}{P_0} \quad (1.1)$$

Діапазон зміни рівнів звуків мови називають динамічним діапазоном мови. Вухо людини не відчуває звуків нижче 0 дБ, а поріг відчуття болю лежить у межах 120...125 дБ.

Рівні чутності звуків виражають у фонах та позначають літерою L . Рівень чутності L визначає число децибел рівня тону з частотою 1000 Гц, який рівнозначний зі звуком, що досліджується. Він також залежить від частоти (рис. 1.3). Виміри рівня чутності виконують за допомогою шумоміра.

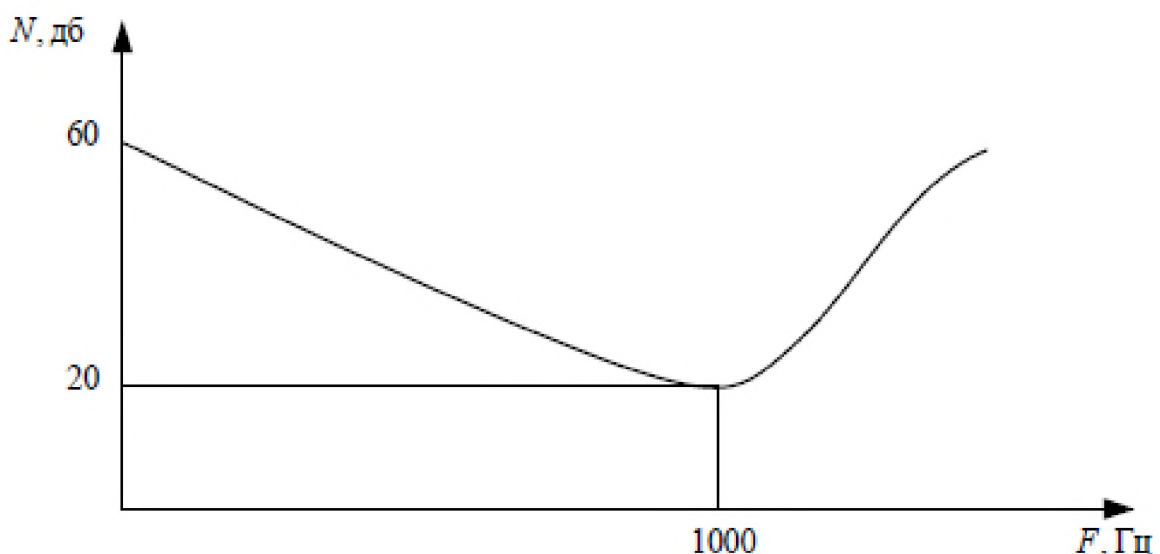


Рисунок 1.3 – Залежність гучності від частоти

Рівень різних частотних складових мови неоднаковий. Залежність усереднених за великий термін часу рівнів мови від частоти називають спектром мови. На рис. 1.4 наведено спектр української мови, звідки видно, що найбільші рівні мають складові мови в діапазоні 200...600 Гц, і тут зосереджена основна частка енергії мови.

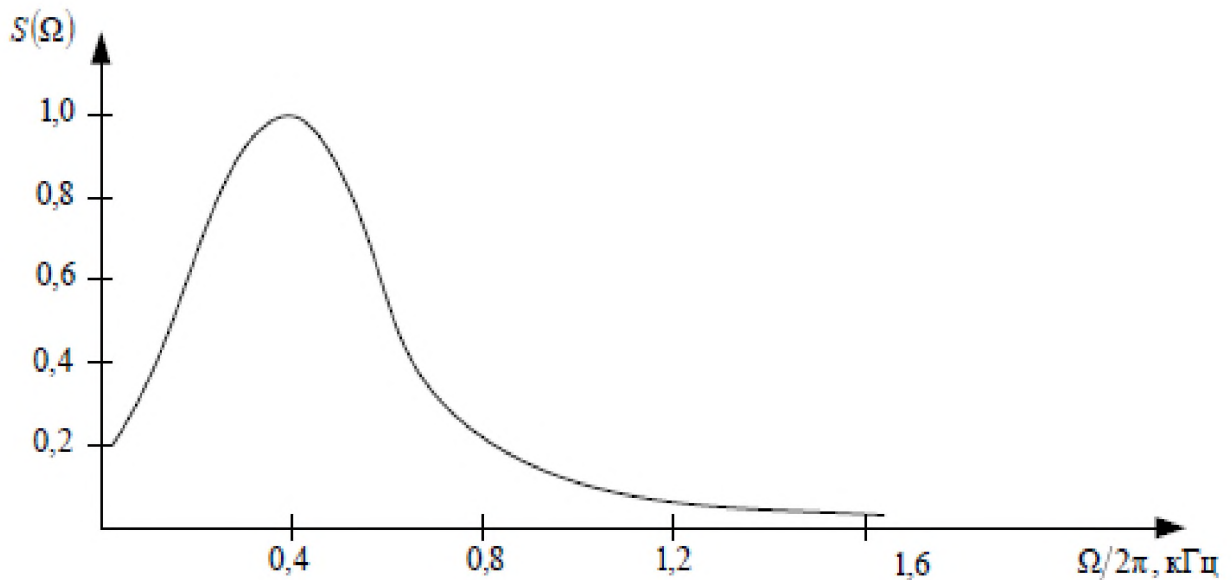


Рисунок 1.4 – Спектр української мови

Цікавою особливістю спектра мови є те, що в цьому частотному діапазоні можна отримати тільки інформацію про частоту основного тону та про першу форманту. Тому, у разі аналізу слухового апарату людини слід звернути увагу на те, яким чином компенсується такий невдалий розподіл енергії в спектрі мови.

Досить часто в літературі доводиться зустрічатися з такими термінами, як пікфактор та рівень спектра звуку. Під пікфактором мови (P_p) розуміють відношення пікового тиску до середнього ефективного:

$$P_p = 20 \lg \frac{P_{\text{пик}}}{P} \quad (1.2)$$

Рівень спектра звуку (B) є тим же рівнем інтенсивності звуку, приведеним до 1 Гц:

$$B = 10 \lg \frac{I}{I_0 \Delta f_E} \quad (1.3)$$

де f_E – значення еквівалентної смуги частот, в якій вимірювали рівень інтенсивності.

Для пояснення процесу слухового сприйняття мови користуються так званою резонансною теорією слуху. Відповідно до цієї теорії модель вуха уявляють у вигляді набору резонаторів, які настроєні на різні частоти та мають смугу пропускання, що зветься критичними смугами слуху для тієї чи іншої частоти. Вуха здатне інтегрувати частотні складові звукового спектра, які знаходяться у межах критичної смуги. Критична смуга слуху залежить від частоти (рис. 1.5).

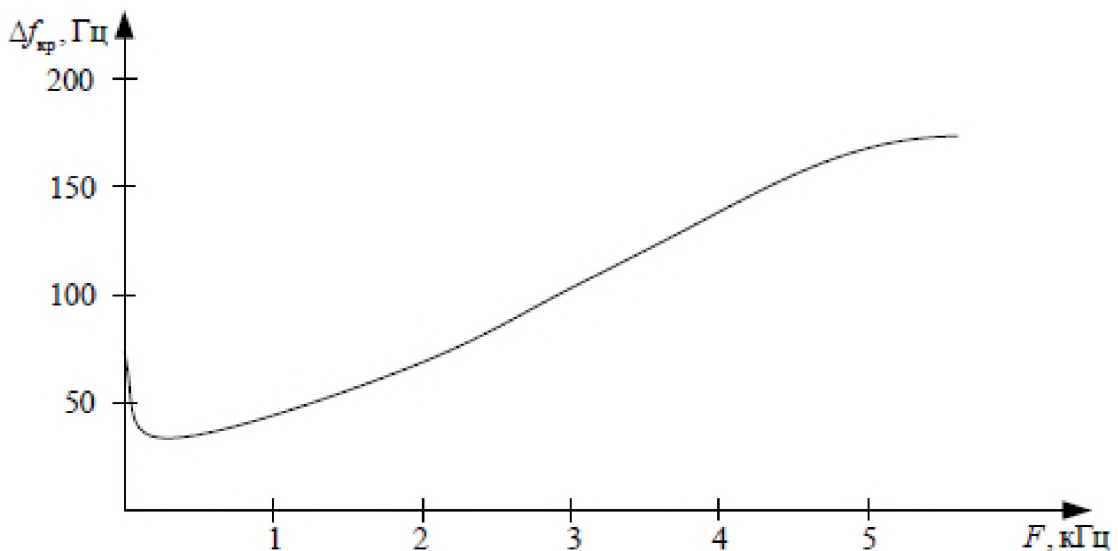


Рисунок 1.5 – Залежність критичної смуги слуху від частоти

Існує деякий мінімальний рівень звуку, який сприймається вухом. Він зветься порогом чутливості, а також є функцією частоти. Так, звуки на різних частотах вуха людини починає чути з неоднаковим рівнем.

На рис. 1.6 показано залежність порога чутливості від частоти. З рис. 1.6 видно, що звуки з частотою 200 Гц починають бути чутними при рівні інтенсивності 30 дБ. В той же час для звуків на частоті 3000 Гц досить навіть 0 дБ, щоб бути почутими вухом людини.

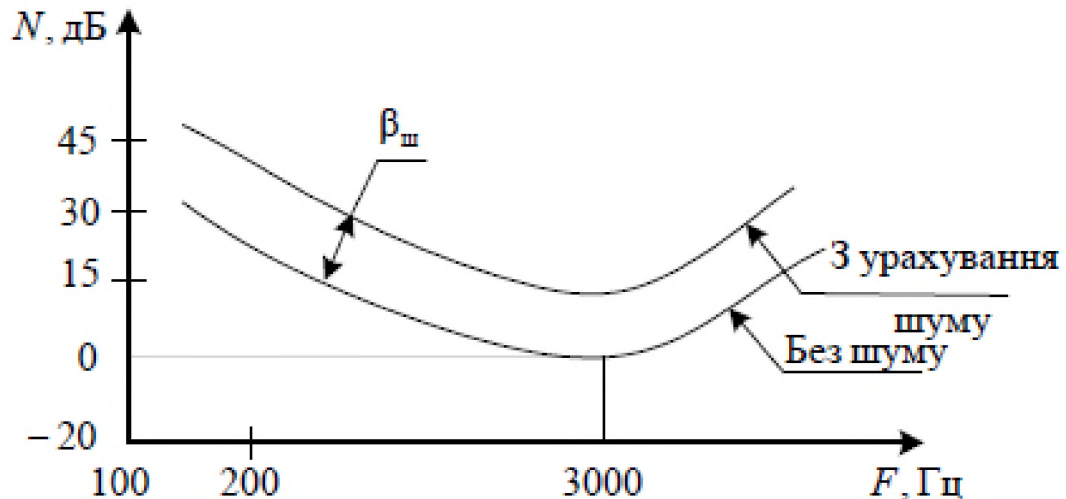


Рисунок 1.6 – Залежність порога чутливості від частоти

Таким чином, вухо людини, немовби, компенсує недоліки розподілу енергії в спектрі мови. Тобто звуки мови, які мають багато енергії, починають бути чутними при високому рівні інтенсивності (звуки в смузі 200...600 Гц), а звуки з малою енергією – відповідно чутні з низьким рівнем (звуки в смузі від 1000 до 4000 Гц).

Зрозуміло, що поріг чутності в тиші відрізняється від порога під час дії шуму на відповідний рівень інтенсивності шуму, який оцінюють як

$$\beta_{\text{ш}} = 10 \lg \frac{I_{\text{ш}}}{I_0}.$$

Слід відзначити, що шум стає відчутним, коли його рівень перевищує 80 дБ. Рівень шуму в 90 дБ потребує вести розмову з напругою, а при рівні в 115 дБ розмова взагалі не можлива.

1.2 Компресія мовних сигналів

1.2.1. Елементи теорії інформації в застосуванні до мовного сигналу

Кількісною мірою мовного повідомлення є кількість відомостей, що містяться в ньому [1]. Так, наприклад, у телеграфному повідомленні кількість відомостей може бути легко визначена з кількості відомостей, що містяться в

кожному знаці. Під час аналізу мовних повідомлень на перший погляд здається, що мову, як неперервний процес, можна представити тільки нескінченно великим числом елементів. Однак із теореми Котельникова відомо, що сигнал, у який перетворюється будь-яке неперервне повідомлення, можна представити з наперед заданою точністю кінцевим числом елементів, тобто передаються тільки ординати безупинного сигналу, розташовані на інтервалах $1/2F_B c$, де F_B – верхня частота діапазону.

У системах авіаційного радіозв'язку – це звичайно сигнали обмеженої тривалості T і смугою спектра F . Процес передачі повідомлення в цьому випадку цілком визначається числом параметрів, рівним [1, 8, 11]

$$N=2FT. \quad (1.4)$$

Цими параметрами можуть бути або миттєві значення сигналу, розташовані на інтервалах $1/2Fc$, або спектральні коефіцієнти розкладання в ряд Фур'є та інші.

Отже, безупинний сигнал кінцевої тривалості і з обмеженим спектром точно описується кінцевим числом значень цього процесу.

Крім квантування в часі, може бути квантування за рівнем (рівнями) миттєвих значень сигналу, при якому відбувається викривлення форми сигналу, однак ці викривлення не перевищують значень викривлень, викликаних завадою. Можна вважати, що ніякого викривлення сигналу під час квантування за рівнем не відбувається, а лише з'являється завада зі ступінчатою зміною свого рівня.

Під час розгляду процесу квантування за рівнем необхідно запам'ятати, що поняття рівня миттєвого значення сигналу не слід змішувати з поняттям рівня потужності сигналу, що часто скорочено називають також рівнем сигналу.

Таким чином, сигнал можна замінити послідовністю імпульсів, розташованих на дискретних інтервалах і дискретних значеннях рівнів. У будь-якому вигляді мовного повідомлення (звук, склад, слово і таке інше) кількість відомостей буде тим більше, чим більше загальне число знаків, з яких даний

знак обраний. Якщо всіх можливих знаків m , а число знаків у повідомленні n , то число можливих сполучень буде $N=m^n$, звідси число знаків у повідомленні

$$n = \frac{\log_2 N}{\log_2 m} \quad (1.5)$$

і максимальне число відомостей (інформації) пропорційне числу знаків у повідомленні:

$$I = \log_2 N = n \log_2 m. \quad (1.6)$$

Максимальна кількість відомостей, що приходяться на один знак повідомлення, тобто змістовність повідомлення, оцінюють у бінарних одиницях (біт):

$$I = \log_2 m. \quad (1.7)$$

Розглянемо взаємозв'язок між основними параметрами сигналу і кількістю відомостей, що містяться в ньому. Як відомо, сигнал визначається тривалістю, частотним і динамічним діапазонами. Нижня границя динамічного діапазону і точність його передачі визначаються рівнем завад. Звичайно вважають, що динамічний діапазон сигналу характеризується відношенням середніх потужностей сигналу P_c і завади P_n чи різницею їхніх рівнів:

$$D = \log \frac{P_c}{P_n} = \log P_c - \log P_n. \quad (1.8)$$

Добуток трьох вимірів сигналу називають обсягом сигналу і записують у такий спосіб:

$$V = FTD \quad (1.9)$$

чи з урахуванням виразу (1.14)

$$V = FT \log \frac{P_c}{P_n}. \quad (1.10)$$

Вираз для кількості відомостей, вважаючи, що кількість знаків у

неперервному сигналі $n=2FT$ і $m^2 = a \frac{P_c}{P_n}$, запишемо в такий спосіб:

$$I = FT \log a \frac{P_c}{P_{\Pi}}, \quad (1.11)$$

де a визначається кодом сигналу.

Таким чином, порівнюючи вирази (1.10) і (1.11), знаходимо, що кількість відомостей і обсяг сигналу розрізняються тільки на постійну величину $\log a$. Якщо врахувати, що кількість відомостей у заданому повідомленні – величина незмінна, то, змінюючи основу коду m , можна відповідно змінювати число елементів сигналу n . Це означає, що, наприклад, змінюючи число переданих рівнів, можна відповідно змінювати ширину частотного діапазону ($n=T/\Delta t=2FT$). Таким чином, шляхом кодування можна змінювати кожний з трьох вимірів сигналу за рахунок іншого (інших). У цьому випадку обсяг сигналу також залишається незмінним.

В процесі передачі інформації з каналу зв'язку в реальному масштабі часу лише час повідомлення T залишається незмінним, смуга ж частот F і відношення P_c/P_{Π} зменшуються. Таким чином, обсяг прийнятого сигналу завжди менше обсягу переданого. Для більш повного використання ліній зв'язку та їхнього спрощення обсяг переданого сигналу бажано зменшувати, однак це часто суперечить вимогам підвищення завадостійкості передачі. У таких випадках необхідно шукати компромісне рішення, обумовлене призначенням лінії зв'язку. При цьому необхідно мати на увазі, що відношення P_c/P_{Π} у виразі (1.11) є інтегральним, і йому відповідає нескінченність комбінацій спектральних відношень P_c/P_{Π} у кожній точці площини TF , і тільки цілком визначені сполучення відношень дають найбільшу завадостійкість передачі при даному відношенні P_c/P_{Π} . Отже, при передачі мови має значення не тільки обсяг сигналу, але і його конфігурація. Тому перетворення обсягу сигналу, вироблене для підвищення ефективності роботи лінії зв'язку, може бути здійснено всілякими способами.

Дослідження показали, що обсяг інформації, перенесеної мовним сигналом, виявляється майже на порядок менше обсягу мовного сигналу. Тому зменшення обсягу сигналу не призводить до скорочення інформації,

перенесеної цим сигналом. Зменшення обсягу мовного сигналу чи його стиснення (компресія) досягається за допомогою відповідних перетворень мовного сигналу. Розрізняють методи безпосередньої і параметричної компресій (компресія з функціональним перетворенням мови). Класифікація методів компресії мови показана на рис.1.7.

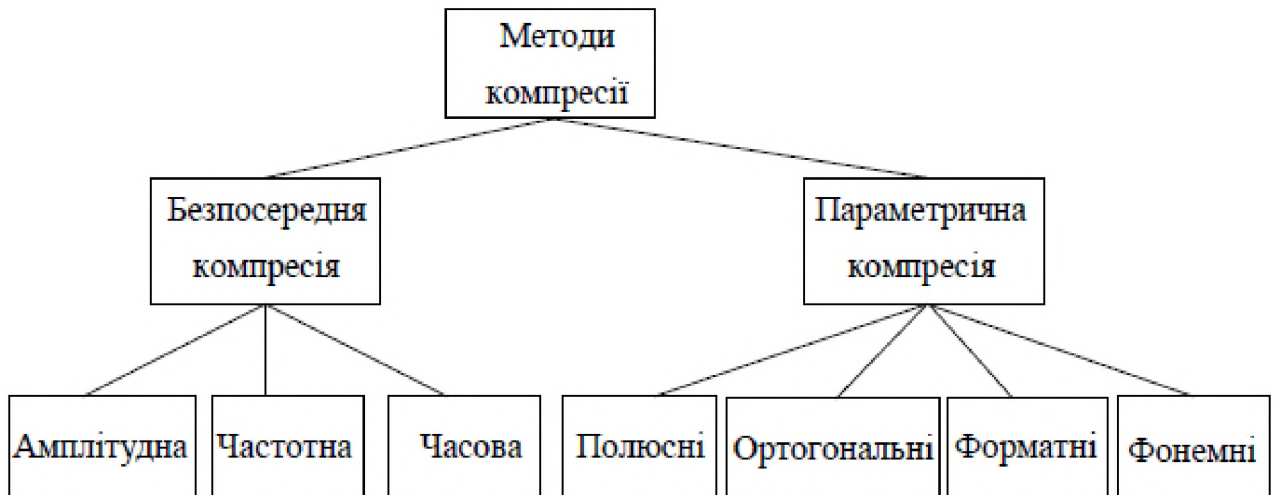


Рисунок 1.7 – Класифікація методів компресії мови

1.2.2 Методи безпосередньої компресії

Компресією мовних сигналів називається зменшення обсягу сигналів за рахунок стиснення одного чи декількох параметрів сигналу (динамічний діапазон, спектр частот, тривалість). Ці методи відрізняються тим, що виробляються тільки деформації обсягу сигналу, а мікроструктура сигналу цілком не знищується. Під час деформації об'єму сигнал частково викривляється, однак ці викривлення відіграють роль завад. Тут сигнал у відповідному вимірі «деформується», тобто стискається на передавальному кінці з відповідним розширенням його на приймальному кінці [1].

Отже, після обмеження сигналу по динамічному, частотному діапазону чи в часі відновити сигнал цілком на приймальному кінці не вдається.

Розгляд методів безпосередньої компресії почнемо з амплітудної компресії. У динамічному діапазоні сигналу міститься переважно інформація про якість звучання й у значно меншому ступені – інформація про розбірливість звуків, тобто компресований за рівнем мовний сигнал має розбірливість, що мало відрізняється від вихідного сигналу. Таким чином, компресія дозволяє підвищувати завадостійкість передачі.

Стиснення динамічного діапазону мови виробляється або методом автоматичного регулювання фонетичного рівня мови, або миттєвою компресією сигналу. При першому способі стиснення рівні звуків мови так чи інакше зближуються один з одним. У результаті піковий і мінімальний рівні компресованої мови також зближуються. Перехідні процеси, що залежать від сталої часу компресуючого пристрою, дуже спотворюють деякі звуки мови (б, п, д, т). В цілому динамічне компандування мови, крім корисного ефекту, має недолік – поява викривлень через наявність перехідних процесів. Корисна дія компандування виявляється в тому, що при заданій піковій потужності передавача, незважаючи на викривлення, розбірливість мови підвищується навіть при наявності завад. У цьому випадку максимальний рівень мови на виході компресора буде таким же, як і на вході, а більш низькі рівні будуть піднятими.

Внаслідок «підтягування» слабких складових мови збільшується її розбірливість на тлі завад. Крім того, стиснення динамічного діапазону при заданій піковій потужності генератора забезпечує краще використання вихідних каскадів. Більш ефективним є спосіб стиснення динамічного діапазону шляхом миттєвої компресії (амплітудного обмеження). Цей спосіб є без інерційним. Розрізняють обмеження по максимуму (рис.1.8, а, в) і по мінімуму (рис.1.8, б, г).

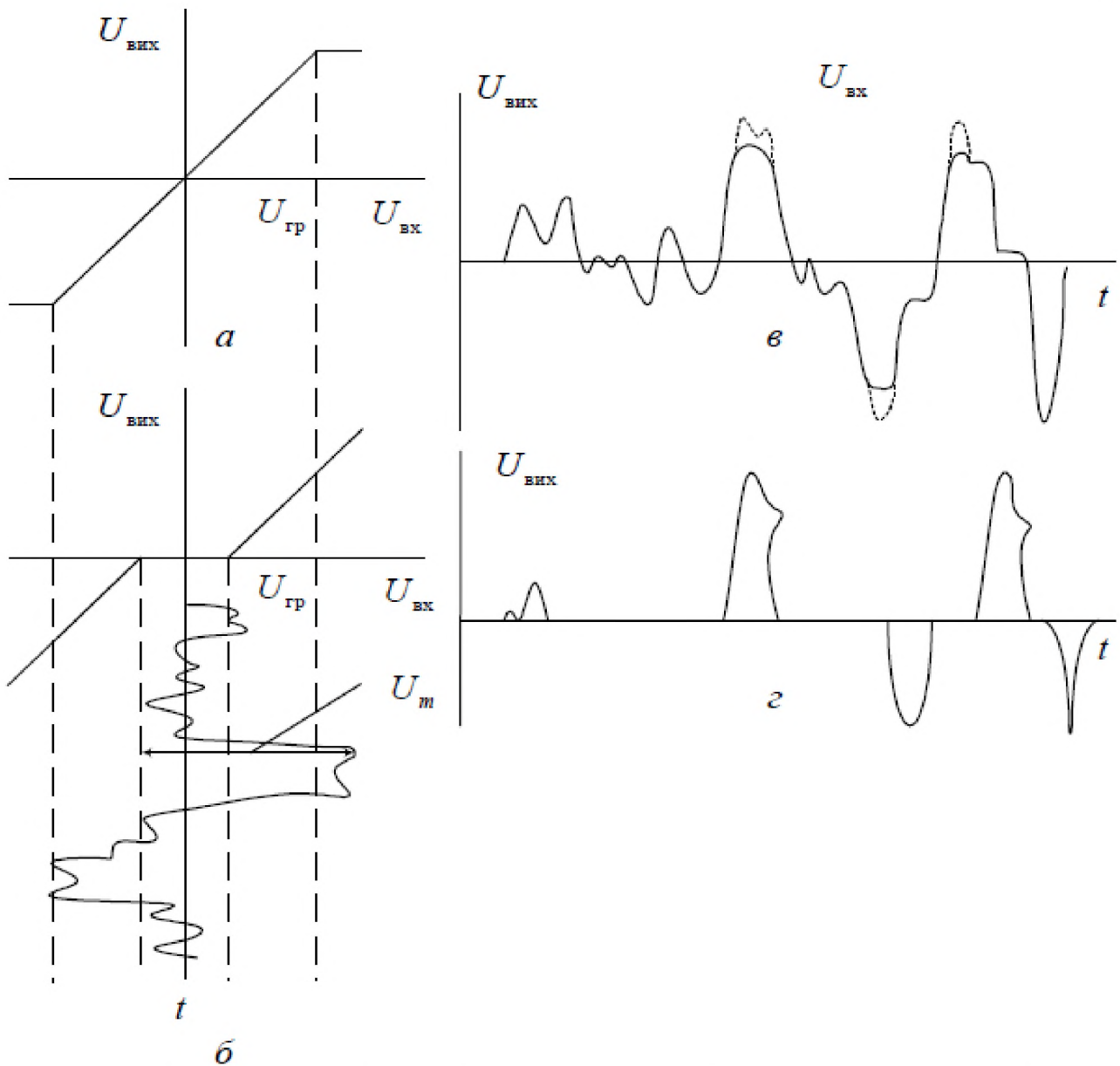


Рисунок 1.8 – Епюри напруг, що пояснюють принцип обмеження по максимуму і мінімуму

Під ступенем обмеження зверху розуміють величину $\gamma_{\text{В}} = 20 \lg \frac{U_{\text{max}}}{U_{\text{обм}}}$, а

при обмеженні знизу – $\gamma_{\text{Н}} = 20 \lg \frac{U_{\text{max}}}{U_{\text{max}} - U_{\text{обм}}}$, де U_{max} – максимальне значення напруги на вході обмежувача; $U_{\text{обм}}$ – напруга початку обмеження.

Вплив обмеження знизу і зверху на розбірливість мови різний (рис. 1.9). Незначне обмеження знизу викликає істотне зниження розбірливості мови, а навіть при сильному обмеженні зверху розбірливість мови неістотно падає.

Встановлено, що висока розбірливість мови зберігається навіть при граничному обмеженні, що називається кліпуванням. Кліпована мова має вигляд прямокутних імпульсів різної тривалості (рис. 1.10). При цьому єдиною інформацією про первісну мову є послідовність нульових переходів. Виходячи з високої розбірливості процесу кліпування, впливає, що ті чи інші властивості розташування нулів (рис. 1.10) містять велику кількість інформації про мовне повідомлення.

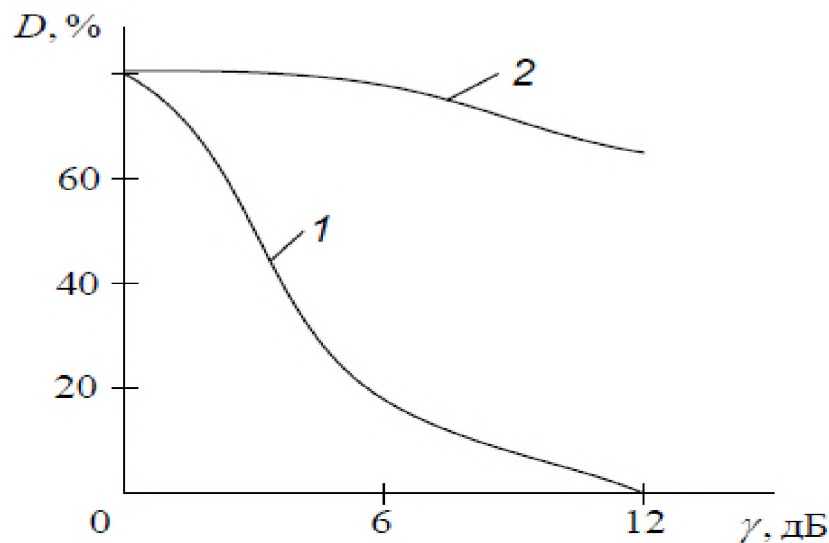


Рисунок 1.9 – Вплив на розбірливість обмеження зверху (2), і знизу (1)

Ще краща розбірливість, якщо мову попередньо диференціюють, а потім обмежують похідну мовного процесу. У цьому випадку зберігається положення не нульових, а екстремальних значень мовного повідомлення. Більш висока розбірливість мови для цього випадку пояснюється тим, що число екстремальних значень у мовному повідомленні більше числа нульових значень. Експериментально встановлено, що для чоловічих голосів нульові значення впливають з частотою 2780 Гц, а екстремальні – з частотою 4700 Гц. Таким чином, при збереженні положень екстремальних значень за допомогою диференціювання зберігається і передається в лінію зв'язку більш повна інформація про структуру сигналу.

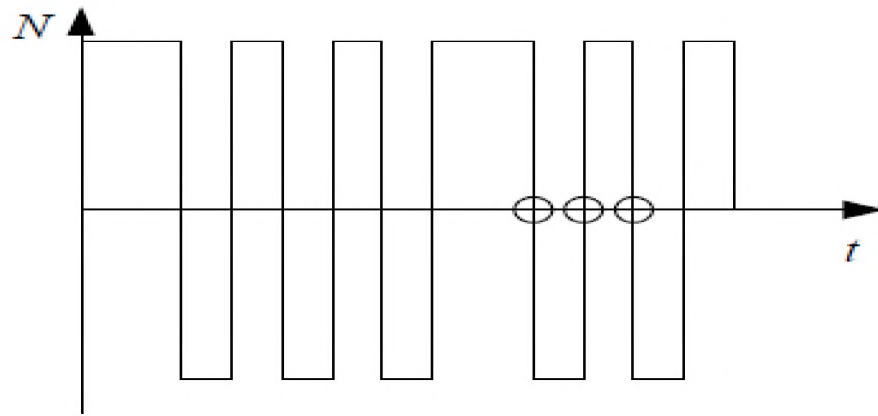


Рисунок 1.10 – Осцилограма кліпованої мови

Мовний сигнал у разі дотримання деяких умов може бути записаний як

$$F(t) = U_F(t) \cos \varphi(t). \quad (1.12)$$

При цьому функції $U_F(t)$ і $\varphi(t)$ визначені у такий спосіб:

$$U_F(t) = \sqrt{F^2(t) + V^2(t)}; \quad \varphi(t) = \operatorname{arctg} \frac{V(t)}{F(t)}, \quad (1.13)$$

де вихідна функція $F(t)$ і сполучена функція $V(t)$ однозначно пов'язані між собою інтегральними перетвореннями Гільберта:

$$V(t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{F(\tau)}{\tau - t} d\tau; \quad F(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{V(\tau)}{\tau - t} d\tau. \quad (1.14)$$

Визначені таким способом функції $F(t)$ і $V(t)$ являють собою на комплексній площині функцію $F(t) + jV(t) = U_F(t)e^{j\varphi(t)}$, що називається аналітичною формою сигналу, утвореного з функції $F(t)$.

Сигнал $F(t)$ є дійсною частиною аналітичного виразу $F(t) = \operatorname{Re} U_F(t)e^{j\varphi(t)}$.

Вхідні у виразі (1.13) функції $U_F(t)$ і $\varphi(t)$ являють собою огибающую і фазу сигналу $F(t)$ і містять відповідно амплітудну і частотну інформації. Очевидно,

що похідна миттєвої фази $\omega(t) = \frac{d\varphi(t)}{dt}$ є миттєва частота сигналу $F(t)$.

Кліпування мови зводиться до усунення амплітудної інформації і до виділення та передачі частотної інформації, що міститься у функції $\cos \varphi(t)$, тобто в передачі мови постійного рівня. Виділення частотної інформації не

можна здійснювати шляхом безмежного стискання, тому що спектри $U_F(t)$ і $\cos \varphi(t)$ перекриваються, а ширина спектра перевершує октаву. У такому випадку частина гармонік, що виникли в результаті сильного обмеження, потрапить у вхідний діапазон частот і буде засмічувати обмежений сигнал на виході. Один зі способів усунення зазначеного недоліку полягає в зменшенні продуктів нелінійного викривлення шляхом попередньої корекції частотної характеристики мови до обмеження.

При стисненні динамічного діапазону сигналу найбільш обмежуються низькочастотні складові спектра, що мають найбільшу частку енергії мови, в результаті чого утворюються інтенсивні гармоніки, що не виходять за межі спектра мови, і тому спотворюють її. Крім того, в обмежувачі має місце безпосереднє придушення високочастотних складових низькочастотними. Якщо перед обмежувачем підняти високочастотні складові мови, то воно буде більш рівномірним, і зазначені викривлення зменшаться. Таке піднімання частотної характеристики може бути отримано під час проходження мовного сигналу через ланку, що диференціює. Це також є причиною підвищення розбірливості попередньо диференційованої обмеженої мови. Ще більш ефективним способом зменшення продуктів нелінійних викривлень є перенесення обмеження в область високих частот. Так, якщо

$$F(t) = U_F(t) \cos[\omega t + \varphi(t)], \quad (1.15)$$

то частоту ω можна вибрати настільки великою, що спектри функцій $U_F(t)$ і $\cos(\omega t + \varphi(t))$ не будуть перекриватися, тобто $U_F(t)$ порівняно з $\cos(\omega t + \varphi(t))$ буде змінюватися повільно. Крім того, можна домогтися, щоб ширина спектра сигналу, зміщеного на частоту ω , була меншою порівняно зі значенням $\omega/2\pi$, тобто вже октави.

Таким чином, гармоніки, що виникають при обмеженнях, виявляються поза межами діапазону частот функції $F(t)$ і тому можуть бути відфільтровані на виході обмежувача.

У результаті замість сигналу прямокутної форми процес на виході фільтра буде мати вигляд частотно-модульованого коливання.

Отже, оптимальний ступінь обмеження залежить від виду модуляції, від заходів, прийнятих для боротьби з нелінійними викривленнями, а також від умов, за яких здійснюється передача мовних повідомлень. Так, під час передачі з тиші оптимальним є обмеження на 18...24 дБ щодо пікового рівня мови. Таке обмеження прийнятне доти, поки відношення шум/сигнал менше одиниці. У разі більш інтенсивних завад розбірливість обмеженого сигналу різко знижується. Розбірливість необмеженого сигналу за таких умов знижується менш різко, однак вона буде також незадовільною. У разі ступеня обмеження 40 дБ розбірливість обмеженого сигналу при будь-якому рівні завад буде нижче розбірливості необмеженого сигналу.

За умов роботи з високим рівнем шуму ($\beta_{\text{ш}}=110...120$ дБ) за рахунок мимовільного збільшення рівня мови оператором на 6...8 дБ ступінь обмеження можна вибирати в межах 18...24 дБ. Слід зазначити, що передача обмеженої мови пов'язана з необхідністю великого підсилення в низькочастотному тракту передавального пристрою. Це призводить до відносного зростання шумів у паузах, що погіршує якість звучання, і є істотним недоліком систем зв'язку, що використовують граничну компресію динамічного діапазону.

Одним зі способів зменшення шумів у паузах є запирання мовного тракту за допомогою обмежувача по мінімуму. Однак у цьому випадку погіршується якість відтворення внаслідок пропадання коротких вибухових звуків, особливо на початку слів. Високу розбірливість і гарну якість звучання мови можна одержати шляхом передачі по рівнобіжних каналах кліпованого сигналу і його огинаючої з наступним перемножуванням на приймальному кінці. Порівнюючи між собою перший і другий способи стиснення динамічного діапазону, бачимо, що при другому способі відбувається більше звуження динамічного діапазону, однак нелінійні викривлення при цьому більш значні. Пристрої, що реалізують цей спосіб, дуже прості, тоді як пристрої автоматичного регулювання рівня досить складні, нестійкі в роботі і вносять помітні викривлення внаслідок процесів, що відбуваються в них.

Існує третій спосіб стиснення динамічного діапазону, що називається амплітудною селекцією. Сутність його полягає в передачі тільки максимальних і мінімальних значень мовного сигналу. Амплітудна селекція, маючи всі достоїнства кліпованої мови, має перевагу перед нею, тому що потребує смугу передачі в 2,5 рази вужчу.

Розглянемо основні методи частотної компресії. Спектр мови займає смугу частот 100...9000 Гц. Найпростіший і найбільш поширений спосіб частотної компресії полягає в обмеженні спектра мовних сигналів, переданих по лінії зв'язку. Для передачі вибирається ділянка спектра, що є найбільш важливою для забезпечення необхідної якості зв'язку. Високочастотні складові мови понад 3...5 кГц несуттєво впливають на її розбірливість, тому з метою звуження спектра телефонного каналу і зниження внаслідок цього рівня завад ці складові доцільно обрізати. Це також доцільно і для низькочастотних складових нижче 250...300 Гц, що мають досить високу інтенсивність, однак мало впливають на розбірливість мови. Результати досліджень показали, що під час телефонної передачі мови, особливо за умов флуктуаційних шумів з рівномірною щільністю по частоті, обмеження частотного діапазону зверху частотою 3500 Гц і знизу частотою 300 Гц призводить лише до підвищення розбірливості мови.

Компресія спектра телефонного сигналу може бути отримана також за рахунок збільшення часу передачі. Якщо мову спочатку записати на плівку і зменшувати частоту обертання магнітофона проти нормальної в n разів, то у таке ж число разів звужиться спектр мови. На приймальній стороні для відновлення нормальної мови частота обертання повинна бути збільшена в n разів. Хоча розглянуті вище методи безпосередньої частотної компресії мають практичне значення, стиснення спектра мови в цьому випадку невелике і дуже далеке від гранично можливого стиснення.

Теоретична межа можливого стиснення спектра телефонного сигналу може бути встановлена на основі формули Шеннона [1, 11]:

$$C = \frac{I}{T} = \frac{FT \log_2 \left(1 + \frac{P_c}{P_n} \right)}{T} = F \log_2 \left(1 + \frac{P_c}{P_n} \right), \quad (1.16)$$

де F – ширина спектра сигналу; P_c/P_n – відношення потужності сигналу до потужності функціональної завади.

Нормальна розмова ведеться зі швидкістю $V=10$ фонем/с. У разі ідеального кодування кількість двійкових одиниць, необхідних для передачі кожної букви, близька до ентропії повідомлення, тобто $H=2$ біт/фонем. Тоді швидкість передачі при розмові $C=VH=20$ біт/с. Якщо взяти відношення

сигналу до завади в каналі зв'язку рівним $\frac{U_{\max}}{\sqrt{2}\sigma} = 3$, тобто $P_c/P_n=9$, що необхідно для системи зв'язку з АМ, то необхідна смуга частот визначиться з умови

$$F = \frac{VH}{\log_2(1+9)} = \frac{VH}{3,32}. \quad (1.17)$$

Якщо $H=2$ біт/фонем, то $F \approx 6$ Гц, а при $H=5$ біт/фонем $F \approx 15$ Гц. Системи телефонного радіозв'язку з безпосередньою компресією, які застосовуються на практиці, займають значно більшу смугу частот.

Існує також метод часової компресії телефонного сигналу, що полягає в збереженні часу передачі і заснований на надмірності мови, обумовленої її тимчасовими характеристиками. Зазначена надмірність характеризується наявністю повторюваних ділянок у тимчасовій функції. Так, при розгляді спектрограм голосних звуків помітні повторювані ділянки, що впливають одна за одною з частотою основного тону. Наявність основного тону в мові може бути визначена за її осцилограмою (див. рис.1.1). Осцилограма являє собою серію згасаючих коливальних процесів. Інтервал між сусідніми коливальними процесами дорівнює періоду основного тону. Статистика показує, що основний тон чоловічих голосів знаходиться в межах від 70...120 до 150...160 Гц із середньою частотою 120 Гц, для жіночих голосів в межах від 180...220 до 300...330 Гц із середньою частотою 240 Гц.

1.3 Параметрична компресія мовних сигналів

1.3.1 Загальні положення

Методи компресії телефонних сигналів з функціональним перетворенням мови (параметрична компресія) ґрунтуються на заміні мови її параметрами, відомості про які передаються в лінію зв'язку для відновлення повідомлення в приймальному пристрої. Найчастіше вибирають параметри, які повільно змінюються в часі, тому для передачі інформації про такі параметри треба використовувати смугу частот, вузчу ніж для передачі самої мови [1-12].

Пристрої для функціонального перетворення мови називають вокодерами. Робота вокодерів ґрунтується на моделюванні людської мови з урахуванням її характерних особливостей. Замість безпосереднього вимірювання амплітуди вокодер перетворює вхідний сигнал в деякий інший, схожий на первинний. При цьому характеристики мовного сигналу використовують для коригування параметрів прийнятої моделі мовного сигналу. Саме ці параметри і передаються приймачу, який за ними відновлює первинний мовний сигнал. По суті йдеться про синтез мови. У цьому випадку випромінювання спотворень відношення сигнал/шум не має сенсу для вокодерів, а тому необхідні інші суб'єктивні оцінки, такі як середня експертна оцінка, діагностичний римований текст та інші оцінки.

Вокодери можна розділити на два класи: мовноелементні та параметричні. У мовноелементних вокодерах під час передачі розпізнаються елементи мови (наприклад фонемні) і передаються тільки їх номери. На приймальному кінці ці елементи створюють за правилами мовоутворення або беруть із пам'яті пристрою. Галузь застосування фонемних вокодерів – лінії командного зв'язку, мовне керування та інформаційно-довідкові служби. Практично в таких вокодерах здійснюється автоматичне розпізнавання слухових образів, а не визначення параметрів мови [1].

В параметричних вокодерах з мовного сигналу виділяють два типи параметрів:

- параметри, які характеризують огинаючу спектра мовного сигналу (фільтрову функцію);
- параметри, які характеризують джерело мовних коливань (генераторну функцію) – частоту основного тону, її зміни в часі, моменти появи та зникнення основного тону, шумового сигналу.

За цими параметрами на приймальній стороні синтезують мову.

За принципом визначення параметрів фільтрової функції мови розрізняють вокодери:

- смугові каналні (channel);
- формантні;
- ортогональні;
- ліпредери (з лінійним передбаченням мови);
- гомоморфні.

У смугових вокодерах спектр мови ділиться на 7...20 смуг (каналів) аналоговими або цифровими смуговими фільтрами. Велике число каналів у вокодері дозволяє збільшити натуральність та розбірливість. З кожного смугового фільтра сигнал надходить на детектор та фільтр низьких частот з частотою зрізу F_3 . Таким чином, сигнали на виході кожного каналу змінюються з частотою не більше F_3 . Їх передача можлива в аналоговому та цифровому вигляді.

У формантних вокодерах огинаюча спектра мови зображується комбінацією формант (резонансних частот голосового тракту). Основні параметри формант – центральна частота, амплітуда та ширина смуги частот.

В ортогональних вокодерах огинаюча миттєвого спектра розкладається в ряд за вибраною системою ортогональних базисних функцій. Коефіцієнти цього розкладання передаються на приймальну сторону. Найбільше поширення отримали гармонічні вокодери, які використовують розкладання в ряд Фур'є.

Вокодери з лінійним передбаченням (LPC – Linear Prediction Coding) ґрунтуються на оригінальному математичному апараті.

Гомоформна обробка дозволяє розділити генераторну та фільтрову функції, які утворюють мовний сигнал.

Враховуючи складність одержання параметрів генераторної функції, широке застосування отримали напіввокодери (VEV – Voice Excited Vocoder), в яких замість сигналів основного тону та тон-шум використовується смуга мовного сигналу. Смуга частот до 1000 Гц передається по каналу зв'язку в аналоговому або цифровому вигляді. Найбільш відомі напіввокодери VELP (Voice Excited Linear Prediction) та RELP (Residual Excited Linear Prediction).

Вокодери VELP використовують голосове збудження та коефіцієнти лінійного передбачення (КЛП). У вокодерах RELP по вихідному сигналу також обчислюють КЛП.

Якість мови вокодерів є функція від швидкості передачі, продуктивності та затримки обробки. Так, наприклад, низькошвидкісні вокодери звичайно мають більшу затримку та нижчу якість мови ніж високошвидкісні.

У зв'язку з тим, що вокодер використовує канал разом з іншими споживачами або Інтернет з іншими інформаційними потоками, максимальна швидкість повинна бути якомога меншою. Метою сучасних розробок є вокодери зі змінною швидкістю. При цьому використовують фіксовану швидкість для мови та низьку швидкість для фонових шумів. Це досягається за допомогою алгоритмів стискання пауз. У цьому випадку використовують детектор активності мови, який визначає, чи є вхідний сигнал мовою, чи фоновим шумом. Якщо сигнал вважається мовою, він кодується на номінальній фіксованій швидкості, а коли сигнал вважається шумом, він кодується на більш низькій швидкості.

На приймальній стороні відбувається генерація комфортного шуму. Спосіб генерації комфортного шуму повинен бути таким, щоб кодер та декодер залишалися синхронізованими, навіть якщо протягом деякого часу передача

даних не здійснюється. Це дозволяє згладжувати переходи між сегментами активної та неактивної мов.

1.1.2 Смугові вокодери

Типова функціональна схема смугового вокодера показана на рис 1.11.

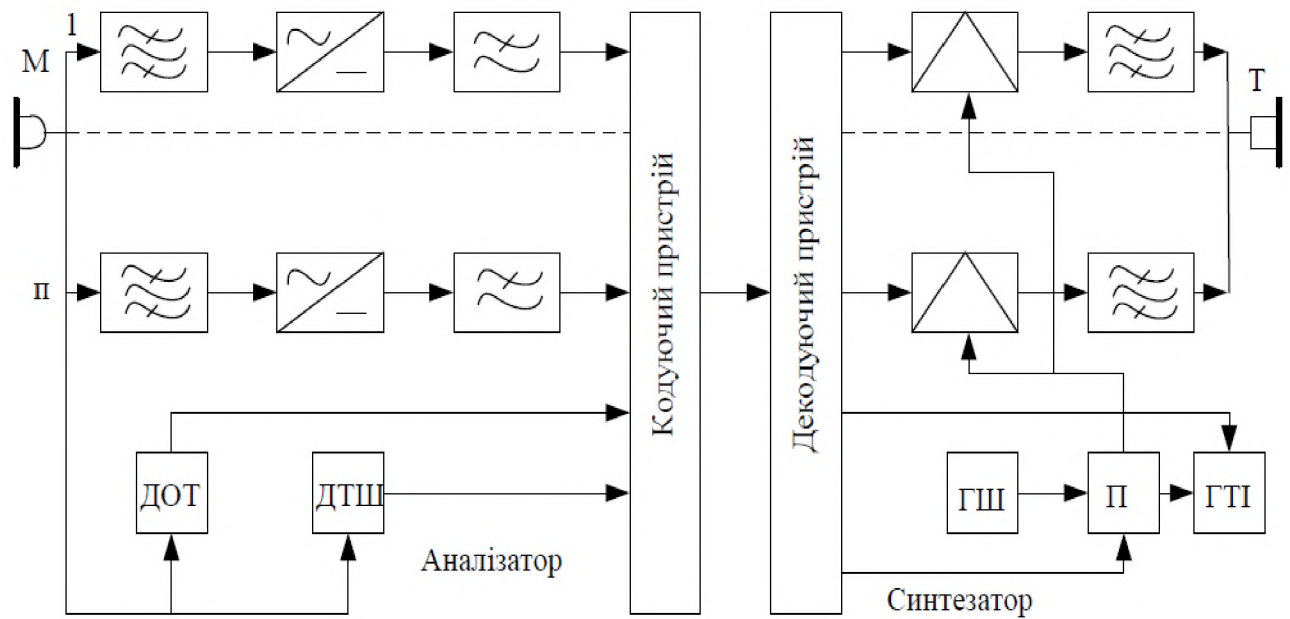


Рисунок 1.11 – Функціональна схема полосного вокодера

Мовний сигнал із мікрофона надходить на гребінку смугових фільтрів (СФ) аналізатора. Кількість фільтрів, а отже і кількість смуг можуть бути різними (від 5 до 20). На виході кожного СФ підключено детектор та згладжувальний фільтр нижніх частот (ФНЧ), який виділяє огинаючу мовного сигналу в даній частотній смузі. Отримана повільно змінювана напруга на виході ФНЧ характеризує амплітуду мовного сигналу в даній смузі частот $a_n(t)$. Практика показує, що при досить великій кількості смуг напруга на виході ФНЧ змінюється повільно, тому за амплітуду мовного сигналу в i -й смузі для інженерних розрахунків приймають значення мовного сигналу на середній частоті i -ї смуги. Мовний сигнал надходить також на пристрій виділення

основного тону (ОТ), на виході якого формується сигнал, який характеризує частоту основного тону $\Omega(t)$.

Крім того, в аналізаторі виділяється сигнал тон-шум (Т-Ш), який характеризує склад спектра звуків мови – дискретний для вокалізованих звуків (тон) або безперервний для невокалізованих звуків (шум). Пристрій виділення сигналів Т-Ш може працювати або безпосередньо від мовних сигналів, або від сигналів, отриманих на виході пристрою виділення ОТ. Тому на рис. 1.11 схема виділення сигналів має два входи.

Сигнали, отримані на виході згладжувального фільтра та на виходах схеми виділення сигналів ОТ та Т-Ш, об'єднуються і перетворюються у форму, яка придатна для передачі каналом зв'язку.

Об'єднання та перетворення сигналів $a_i(t)$, $i = \overline{1, N}$ виконуються в пристрої (рис. 1.11), який називається об'єднуючим. На приймальній стороні каналу зв'язку здійснюється розділення сигналів (у розділювальному пристрої) та перетворення їх у форму, яка необхідна для роботи синтезатора.

В аналогових вокодерах об'єднувальні та розділювальні пристрої реалізуються за принципами частотного розділення сигналів, а в аналогових та цифрових вокодерах – за часовим розділенням сигналів.

Сигнали Т-Ш керують приймачем, за допомогою якого на вхідну гребінку смугових фільтрів подається або широкосмуговий шум від генератора шуму, або імпульси від генератора ОТ. Цей генератор керується сигналом $\Omega(t)$ таким чином, що частота слідування імпульсів на виході дорівнює частоті основного тону мовного сигналу на передавальній стороні. Сукупність генераторів ОТ, шуму та схеми переключення Т-Ш називають генератором мовного сигналу (ГМС).

З виходу смугових фільтрів сигнали надходять на амплітудні модулятори (АМ). На інший вхід АМ як модулюючі надходять сигнали $a_i(t)$, $i = \overline{1, N}$, які після розділювального пристрою проходять через згладжувальні фільтри ФНЧ. З виходу АМ сигнали надходять на гребінку вихідних смугових фільтрів, які застосовуються для зменшення впливу побічних продуктів модуляції, які

виникають в АМ. Сукупність схем та вузлів, у яких перетворюють мовні сигнали в межах кожної з частотних смуг (від входу смугового фільтра до виходу вихідного смугового фільтра синтезатора), називають спектральним каналом смугового вокодера. Звичайно, схеми смугових вокодерів доповнюються пристроями лінійного передбачення, що дозволяє створити смугові вокодери з лінійним передбаченням або ЛПК-вокодери. У таких вокодерах використовуються алгоритми лінійного передбачування, за допомогою яких під час аналізу в передавальному пристрої визначаються коефіцієнти передбачення, а в приймальному пристрої на основі цих коефіцієнтів за допомогою рекурсивного цифрового фільтра синтезується еквівалент голосового тракту.

Ідея методу лінійного передбачення заключається в тому, що величина, яка прогнозується у мовному сигналі $\tilde{\lambda}(h)$ на момент випробування h , визначається як лінійно зважена сума попередніх вибірок:

$$\tilde{\lambda}(h) = \sum_{m=1}^P \tilde{\lambda}(h-m) a_m, \quad (1.18)$$

де $\tilde{\lambda}(h-m)$ – мовний сигнал в попередній момент випробувань; $m=1,2,\dots,P$; a_m – коефіцієнти передбачення.

Інтервали часу між моментами вимірювань частотної дискретизації $t_h - t_{h-1} = 1/F_d$. В момент h , коли відомі $\tilde{\lambda}(h-m)$, але й дійсне значення мовного сигналу $\lambda(h)$, можна визначити похибку передбачення

$$e_h = \lambda_h - \tilde{\lambda}_h, \quad (1.19)$$

а потім підібрати коефіцієнти передбачення a_m таким чином, щоб похибка передбачення була мінімальною.

Звичайно, як критерій мінімізації використовують мінімум середньоквадратичної похибки. В цьому випадку треба визначити такі значення

a_m , при яких $\sum_{h=0}^{M-1} e_h^2 = \min$. Задача мінімізації зводиться до рішення системи лінійних рівнянь відносно a_m .

Пристрій для обчислення помилки передбачення представляє собою фільтр, передатна характеристика якого дорівнює передатній характеристиці фільтра, який імітує мовний тракт.

Усереднення помилки передбачення виконується на інтервалі $\Delta t = M t_h$ вибірок, які утворюють кадр (фрейм). Бажано, щоб довжина аналізованого звуку мови була узгоджена з довжиною кадру, але це технічно виконати складно. Тому звичайно приймають $M=100\dots 200$, що при $F_d=8000$ Гц відповідає довжині кадру $\Delta t \approx 12,5\dots 25$ мс.

Для одержання задовільної якості мовних сигналів, які синтезуються, потрібно підрахувати не менше десяти коефіцієнтів передбачення a_m . Враховуючи це, в аналізаторі ЛПК-вокодера треба вирішувати 100...200 лінійних рівнянь з 10...12 невідомими.

Коефіцієнти передбачення, значення яких передаються каналом зв'язку, використовуються як перемінні параметри у рекурсивному цифровому фільтрі, на вхід якого подаються сигнали збудження. Як сигнали збудження в ЛПК-вокодері використовуються такі ж сигнали, які мають місце на виході ГМС у смугових вокодерах. Під час відтворення вокалізованих звуків – це послідовність імпульсів ОТ, а невокалізованих звуків – це випадкова послідовність імпульсів, сформованих генератором шуму.

Замість коефіцієнтів передбачення в більшості варіантів схем ЛПК-вокодерів передбачено отримання еквівалентного набору величин, які носять назву коефіцієнтів відбиття K_0 . Ці параметри менш чутливі до квантування, ніж коефіцієнти передбачення a_m . Набори параметрів a_m та K_0 пов'язані поміж собою набором стандартних рекурентних співвідношень.

Ширина смуги фільтрів на вході та виході вибирається з урахуванням електроакустичних перетворювачів, які використовуються при цьому. Перетворювачі аналог-код та код-аналог працюють на принципах імпульсно-

кової модуляції (ІКМ). Аналізатор сигналів збудження здійснює виділення сигналів ОТ та Т-Ш, а також загального рівня (огинаючої мовного сигналу). Інші вузли виконують ті ж функції, що й у смуговому вокодері без лінійного передбачення.

Під час синтезу та дослідження смугових вокодерів з ЛПК використовують різні моделі мовного процесу. Найбільш точна модель мови представляє собою нестационарний випадковий процес з повільно змінюваною дисперсією та спектральною щільністю. У разі використання такої моделі можна одержати найбільш точний результат оцінки якості вокодера.

1.4 Існуючі підходи до векторного квантування параметрів лінійного передбачення

Системи низькошвидкісного кодування призначені для компактного представлення мовного сигналу, необхідного для передачі по каналах зв'язку з обмеженою пропускну здатністю, без значної втрати його якості. У використовуваному низькошвидкісними системами діапазоні швидкостей (<16 кбіт/с) домінуюче становище серед різних методів кодування займає кодування на основі методу лінійного передбачення (ЛП).

Відмінною ознакою методу ЛП є однокрокове наближене представлення (прогноз) поточного відліку $s(k)$ мовного сигналу лінійною комбінацією $s_{mp}(k)$ кінцевого числа попередніх відліків:

$$s(k) = s_{mp}(k) + G \cdot c(k) = \sum_{i=1}^M b_i \cdot s(k-i) + G \cdot c(k), \quad (1.20)$$

де k – номер поточного відліку; $c(k)$ – породжуючий сигнал з одиничною дисперсією; G – масштабуючий коефіцієнт; M – порядок передбачення; b_i – коефіцієнти передбачення, $i=1,2,\dots,M$.

У системах низькошвидкісного кодування, заснованих на методі ЛП, порядок передбачення M , як правило, є постійним, а значення коефіцієнтів b_i фіксуються на коротких, що примикають один до одного і рівних по тривалості

часових інтервалах, які називаються кадрами мовного сигналу [13]. Аналіз пропонує алгоритмічних рішень низькошвидкісних систем кодування дозволяє виділити домінуючі наразі тенденції в удосконаленні ліпрудерного методу кодування мови:

1) переважне використання для опису поточного стану голосового тракту M -мірного вектора лінійних спектральних частот (ЛСЧ) $\Omega = (\omega_1, \dots, \omega_M)$, що є математичним еквівалентом множини $\{b_i\}$, але який володіє кращою стійкістю до каналних похибок;

2) широке застосування різновидів векторного квантування (ВК) для відображення кодованих параметрів мови при активному використанні внутрішньокадрових і міжкадрових залежностей останніх з метою максимально можливого виключення надмірності.

Алгоритм визначення ЛСЧ відомий і представлений в [14]. Без втрати спільності надалі в роботі під параметрами ЛП розуміється ЛСЧ.

Відомий підхід до векторного квантування параметрів ЛП [14], згідно з яким квантування вектора параметрів ЛП здійснюється в два етапи, причому вектор другого етапу розділяється на два субвектори.

Недоліком відомого підходу до векторного квантування параметрів ЛП CS-ACELP [14] є велике значення похибки квантування, що представляє собою різницю між квантованим і початковим вектором параметрів ЛП. Слід зауважити, що зазначений недолік обумовлений малим числом біт, що відводиться на квантування ЛСЧ в низькошвидкісних системах кодування мовного сигналу.

Відомий підхід до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15], який полягає у тому, що при кодуванні використовують акустико-фонетичну класифікацію оброблюваних кадрів мовного сигналу на чотири непересічних класи, класифікаційне рішення включають в структуру кодової комбінації, переданої по каналу зв'язку, і використовують для детермінування режиму функціонування класифікованого ВК, які відрізняються різними обсягами

кової комбінації (КК) по кожному кодованому параметру, чим забезпечують адаптивний розподіл інформаційних ресурсів.

Відомий підхід до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15] відноситься до електров'язку. Його використання для кодування і декодування мовного сигналу методом лінійного передбачення при адаптивному розподілі інформаційних ресурсів кодека забезпечує досягнення технічного результату у вигляді підвищення якості синтезованого сигналу без збільшення швидкості кодування. Зазначений результат досягається за рахунок використання акустико-фонетичної класифікації оброблюваних кадрів мовного сигналу на чотири непересічних класи – відсутність мови, вокалізована мова, невокалізована мова і перехід до вокалізованої мови. Ця класифікація виконується одночасно із ідентифікацією фільтра короткочасного лінійного передбачення, і за її результатами інформаційні ресурси адаптивно розподіляються по кодованим параметрам. Класифікаційне рішення включають в структуру кодової комбінації для передачі по каналу зв'язку і використовують для визначення режимів векторних квантувача і деквантувача, навчених для кожного класу мовних кадрів.

Такий підхід підвищує ступінь адаптації процедур кодування та декодування від параметричної до структурної і дозволяє використовувати відмінності в статистичних характеристиках кодованих параметрів мовного сигналу в зазначених класах мовних кадрів, виділяючи біти на кодування параметрів пропорційно їх інформативності в даному класі мовних кадрів. Процедура акустико-фонетичної класифікації виконується одночасно із процедурою ідентифікації моделі, яка настроюється (цифрового полюсного фільтра), що запобігає небажаному зростанню алгоритмічної часової затримки на обробку мовного кадру, що є критичним параметром для здійснення телефонного обміну в режимі реального часу. Класифікаційне рішення h (номер класу оброблюваного мовного кадру) є додатковим параметром

інформаційного обміну між кодером і декодером, в результаті чого кодова комбінація двійкового мультиплікативного коду має вигляд:

$$\vec{A}_{\text{код}} = (\vec{X}_{1\text{код}}, \vec{X}_{2\text{код}}, \vec{X}_{3\text{код}}, \vec{X}_{4\text{код}}, \vec{X}_{5\text{код}}, \vec{X}_{1\text{код}}'', \vec{X}_{2\text{код}}'', \vec{X}_{3\text{код}}'', \vec{X}_{4\text{код}}'', \vec{X}_{5\text{код}}'') \quad (1.21)$$

без зміни її розрядності (без зміни швидкості кодування мовного сигналу). Кодування параметра h вимагає виділення двох біт кодової комбінації, інші інформаційні ресурси адаптивно розподіляються по кодованим параметрам, що обумовлює перехід від використовуваних фіксованих векторних квантування і деквантування кодованих параметрів мовного сигналу до класифікованих векторних квантуванню і деквантуванню із чотирма режимами функціонування. Навчання квантувачів і деквантувачів для кожного класу мовних кадрів проводиться за навчальними вибірками, сформованими на основі мовних кадрів, що належать до даного класу.

На рис. 1.12 наведено пристрій, що призначений для здійснення відомого підходу до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15].

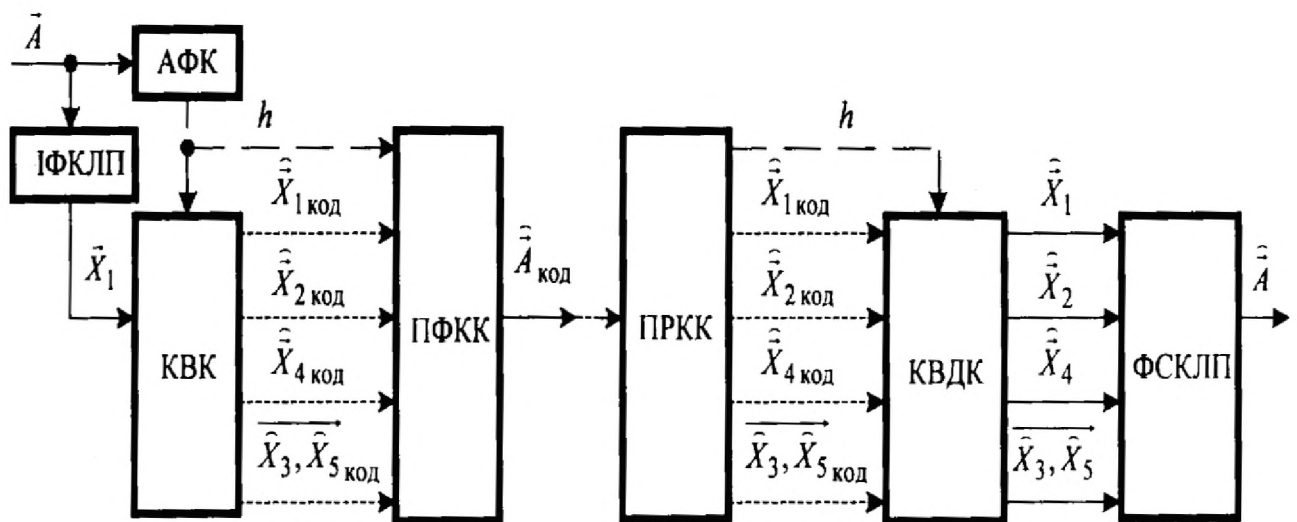


Рисунок 1.12 – Структурна схема пристрою перетворення мовного сигналу на основі методу лінійного передбачення із адаптивним розподілом інформаційних ресурсів, за допомогою якого реалізується відомий підхід [15]

Пристрій (рис. 1.12) складається із передавальної частини (кодера) і приймальної частини (декодера). Кодер містить акустико-фонетичний класифікатор (АФК), ідентифікатор фільтра короткочасного лінійного передбачення (ІФКЛП), класифікований векторний квантувач параметрів мовного сигналу (КВК), пристрій формування кодової комбінації (ПФКК), вихід якого через канал зв'язку з'єднаний з декодером. Декодер пропонованого пристрою містить пристрій розподілу кодової комбінації (ПРКК), класифікований векторний деквантувач параметрів мовного сигналу (КВДК) і фільтр синтезу короткочасного лінійного передбачення (ФСКЛП).

Недоліками відомого підходу до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15] є значна похибка квантування параметрів ЛП, збільшення числа переданих біт, а також артефакти звучання синтезованої мови, викликані перемиканням ВК для кожного класу кадрів мовного сигналу.

Найбільш близьким, прийнятним за прототип, є підхід для векторного квантування з надійним прогнозом параметрів лінійного передбачення в кодуванні мови зі змінною бітовою швидкістю [16], що полягає в наступному. Приймають вхідний вектор параметрів лінійного передбачення, класифікують кадр мовного сигналу, відповідний вхідному вектору параметрів лінійного передбачення, обчислюють вектор передбачення, видаляють обчислений вектор передбачення з вхідного вектора параметрів лінійного передбачення для створення вектора похибки передбачення. Далі масштабують вектор похибки передбачення, квантують масштабований вектор похибки передбачення. При цьому обчислення вектора передбачення включає в себе вибір однієї з безлічі схем передбачення щодо класифікації кадру мовного сигналу і обчислення вектора передбачення відповідно до обраної схеми передбачення. Масштабування вектора похибки передбачення включає в себе вибір щонайменше однієї з безлічі схем масштабування щодо обраної схеми передбачення і масштабування вектора похибки передбачення відповідно до обраної схеми масштабування.

Множина схем передбачення включає в себе передбачення ковзаючим середнім значенням і передбачення авторегресією, при цьому класифікація кадру мовного сигналу включає в себе визначення, що кадр мовного сигналу є стаціонарним вокалізованим кадром. Вибір однієї з безлічі схем передбачення включає в себе вибір передбачення авторегресією, обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення авторегресією. Вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, і масштабування вектора похибки передбачення включає в себе масштабування вектора похибки передбачення із використанням згаданого масштабного коефіцієнта.

Технічний результат використання відомого підходу для векторного квантування з надійним прогнозом параметрів лінійного передбачення в кодуванні мови зі змінною бітовою швидкістю [16] (у порівнянні із попередніми аналогами) – зменшення похибок квантування.

На рис. 1.13 зображений можливий варіант, що не призначений для обмеження, двохетапного векторного квантувача згідно з відомим підходом [16].

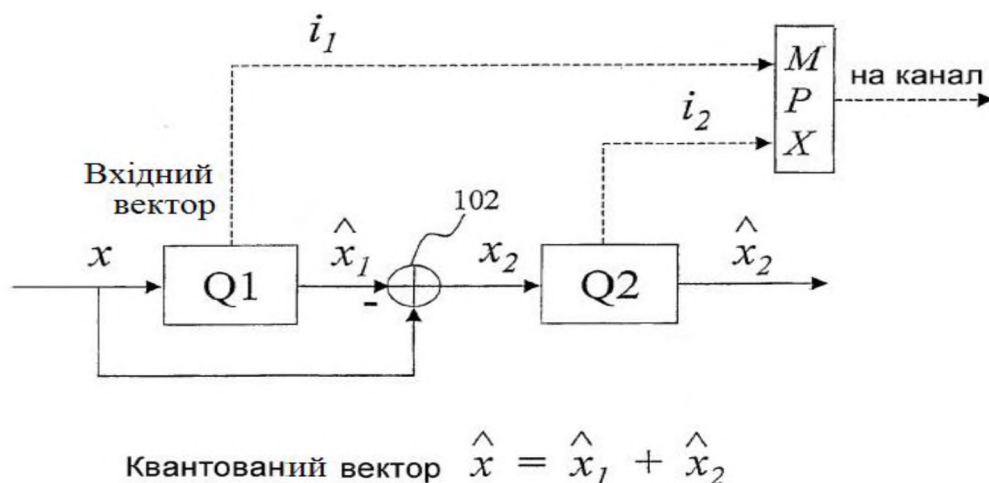


Рисунок 1.13 – Схема, яка ілюструє можливий варіант, що не призначений для обмеження, багатоетапного векторного квантувача згідно з підходом [16]

Спочатку вхідний вектор x квантується квантувачем Q1 для створення квантованого вектору i індексу квантування i_1 (рис. 1.13). Далі обчислюється різниця між вхідним вектором x та квантованим на першому етапі вектором для створення вектора похибки x_2 , який додатково квантується на другому етапі Q2 для створення квантованого вектора похибки другого етапу з індексом квантування i_2 . Індеси i_1 і i_2 передаються через канал, і квантований вектор відтворюється в декодері, як $\hat{X} = \hat{X}_1 + \hat{X}_2$.

На рис. 1.14 зображено ілюстративний можливий варіант векторного квантувача з розподілом.

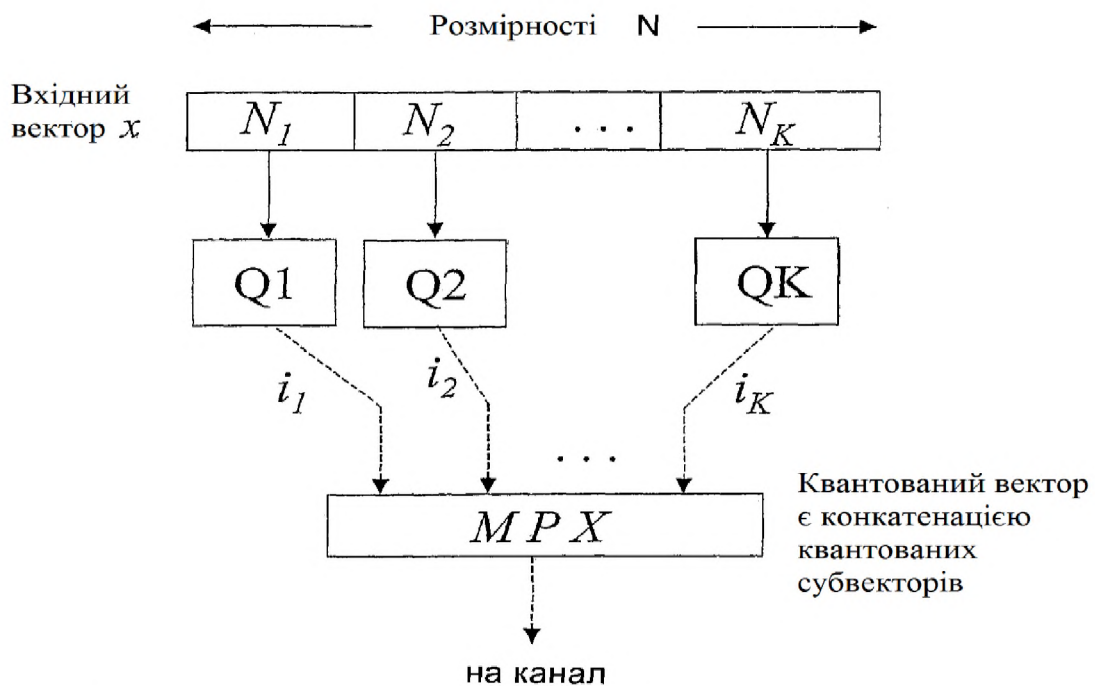


Рисунок 1.14 – Схема, яка ілюструє можливий варіант, що не призначений для обмеження, векторного квантувача з розподілом вектора згідно з підходом [16]

Вхідний вектор x розмірності M поділяється на K субвекторів з розмірностями N_1, N_2, \dots, N_K і квантується векторними квантувачами Q_1, Q_2, \dots, Q_K , відповідно (рис. 1.14). Створюються квантовані субвектори $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_K$ з індексами квантування i_1, i_2 і i_K . Індеси квантування передаються через канал, і квантований вектор відтворюється простою конкатенацією квантованих субвекторів.

Ефективний підхід для векторного квантування повинен об'єднувати багатоетапне Q_1 і Q_2 з розподілом, що призводить до хорошого компромісу між якістю і складністю. У першому ілюстративному можливому варіанті (рис. 1.13) може використовуватися двоетапне Q_1 , при якому вектор похибки другого етапу k_2 розділяється на декілька субвекторів і квантується квантувачами другого етапу $Q_{21}, Q_{22}, \dots, Q_{2K}$, відповідно. У другому ілюстративному можливому варіанті (рис. 1.14) вхідний вектор може бути розділений на два субвектори, потім кожен субвектор квантується двоетапним Q_2 із використанням додаткового розподілу на другому етапі, як і в першому ілюстративному можливому варіанті (рис. 1.13).

Слід зазначити, що зазвичай використовуються два види передбачення: передбачення авторегресією і передбачення ковзаючим середнім значенням. В передбаченні авторегресією передбачений вектор обчислюється як комбінація квантованих векторів з минулих кадрів. У передбаченні ковзаючим середнім значенням передбачений вектор обчислюється як комбінація векторів похибки передбачення з минулих кадрів.

Рис. 1.15 ілюструє можливий варіант векторного квантувача з прогнозом, що використовує передбачення авторегресією.

Як зображено на рис. 1.15, спочатку розраховується вектор похибки передбачення e_n за допомогою вирахування вектора передбачення p_n з вхідного вектора параметрів LP, який повинен квантуватися, x_n . Символ n тут відноситься до індексу кадру в часі. Вектор передбачення p_n обчислюється передбачувачем P з використанням минулих квантованих векторів параметрів LP, $\hat{x}_{n-1}, \hat{x}_{n-2}$, тощо. Далі вектор похибки передбачення e_n квантується для створення індексу і для передачі, наприклад, через канал, і квантованого вектора похибки передбачення k_n . Повний квантований вектор параметрів LP \hat{x}_n отримують за допомогою підсумовування квантованого вектора похибки передбачення k_n і вектора передбачення p_n .

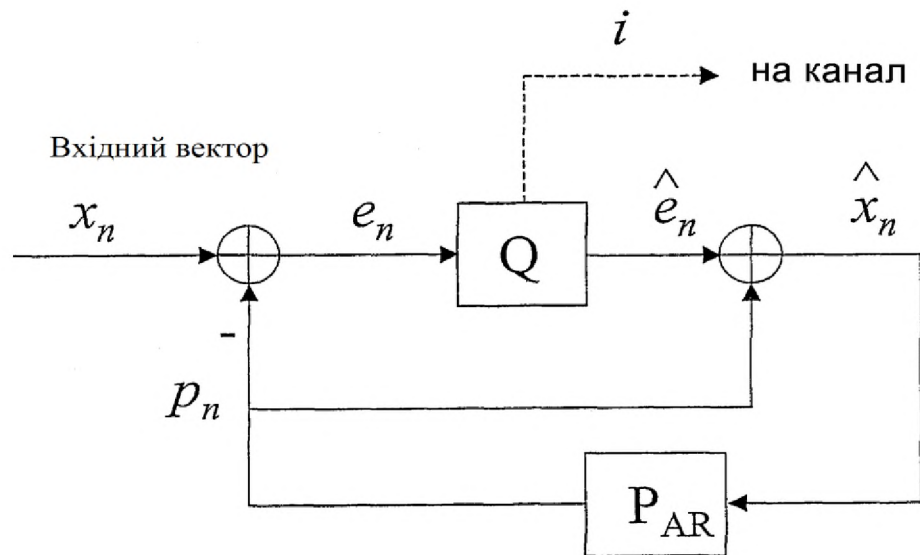


Рисунок 1.15 – Схема, яка ілюструє можливий варіант, що не призначений для обмеження, векторного квантувача з прогнозом, що використовує передбачення авторегресією згідно з підходом [16]

Можливий варіант, що не призначений для обмеження, векторного квантувача прогнозом, що використовує передбачення ковзаючим середнім значенням згідно з підходом [16] зображений на рис. 1.16.

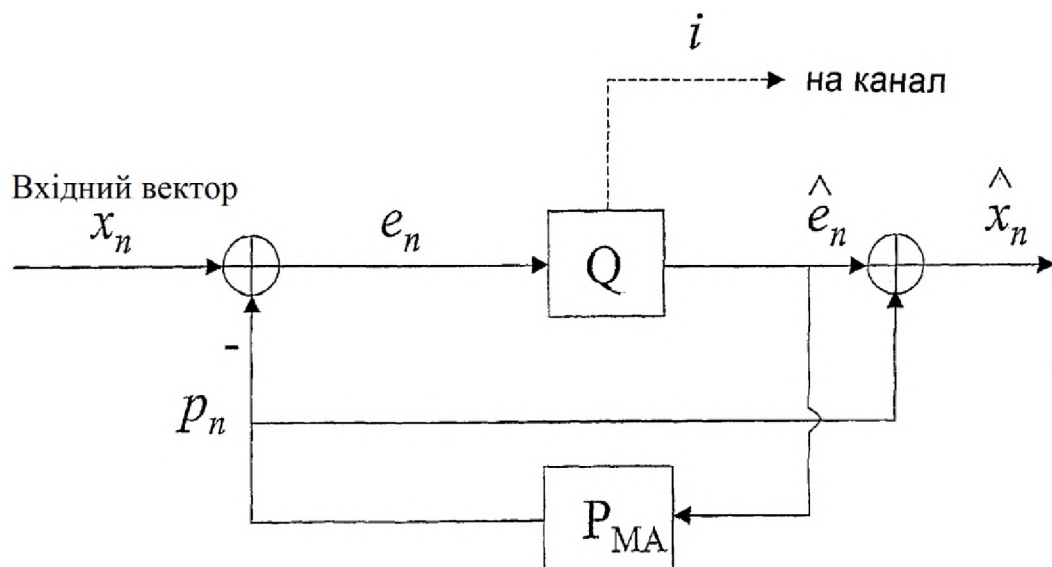


Рисунок 1.16 – Схема, яка ілюструє можливий варіант, що не призначений для обмеження, векторного квантувача з прогнозом, що використовує передбачення ковзаючим середнім значенням згідно з підходом [16]

Як зображено на рис. 1.16, спочатку розраховується вектор похибки передбачення e_n за допомогою вирахування вектора передбачення p_n з вхідного вектора параметрів LP, який повинен квантуватися, x_n . Вектор передбачення p_n обчислюється передбачувачем P із використанням минулих квантованих векторів параметрів LP, \hat{x}_{n-1} , \hat{x}_{n-2} , тощо. Далі вектор похибки передбачення e_n квантується для створення індексу i для передачі, наприклад, через канал, і квантованого вектора похибки передбачення k_n . Повний квантований вектор параметрів LP \hat{x}_n отримують за допомогою підсумовування квантованого вектора похибки передбачення k_n і вектора передбачення p_n .

Не зважаючи на те, що відомий підхід для векторного квантування з надійним прогнозом параметрів лінійного передбачення в кодуванні мови зі змінною бітовою швидкістю [16] має багато можливих варіантів реалізації (рис. 1.11-1.14), він має певні недоліки.

Недоліками відомого підходу [16] є наступні:

1) передбачення авторегресією, що застосовується для стаціонарних вокалізованих кадрів, не є надійним в умовах втрати кадрів – в разі втрати мовних кадрів похибка поширюється на наступні кадри;

2) для отримання мінімальної похибки квантування параметрів ЛП стаціонарних вокалізованих кадрів потрібно витратити значне число біт.

1.5 Висновок. Постановка задачі

В розділі проаналізовано основні положення формування і обробки мовних сигналів. Встановлено, що у зв'язку з активним розвитком техніки зв'язку, рішення задачі кодування мовної інформації має велике значення. Встановлено, що кодування з лінійним передбаченням є одним з найпотужніших методів аналізу мови, і одним з найкорисніших для кодування хорошої якості мови при низькій швидкості передачі бітів. До того ж LPC забезпечує надзвичайно точні оцінки параметрів мови.

В розділі проаналізовано існуючі підходи до векторного квантування параметрів лінійного передбачення. Встановлено, що недоліком відомого підходу до векторного квантування параметрів ЛП CS-ACELP [14] є велике значення похибки квантування, що представляє собою різницю між квантованим і початковим вектором параметрів ЛП. Слід зауважити, що зазначений недолік обумовлений малим числом біт, що відводиться на квантування ЛСЧ в низькошвидкісних системах кодування мовного сигналу..

Встановлено, що недоліками відомого підходу до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15] є значна похибка квантування параметрів ЛП, збільшення числа переданих біт, а також артефакти звучання синтезованої мови, викликані перемиканням ВК для кожного класу кадрів мовного сигналу.

Встановлено, що недоліками відомого підходу для векторного квантування з надійним прогнозом параметрів лінійного передбачення в кодуванні мови зі змінною бітовою швидкістю [16] є наступні:

1) передбачення авторегресією, що застосовується для стаціонарних вокалізованих кадрів, не є надійним в умовах втрати кадрів – в разі втрати мовних кадрів похибка поширюється на наступні кадри;

2) для отримання мінімальної похибки квантування параметрів ЛП стаціонарних вокалізованих кадрів потрібно витратити значне число біт.

Таким чином, для усунення недоліків існуючих підходів необхідно:

- запропонувати підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови;
- оцінити ефективність розробленого підходу.

2 СПЕЦІАЛЬНА ЧАСТИНА

2.1 Підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови

Як вже зазначалось у розділі 1, сучасні методи кодування мовного сигналу можна класифікувати за принципами кодування, які покладені в основу методу. Розрізняють такі принципи кодування мовного сигналу:

- принцип кодування форми хвилі мовного сигналу;
- принцип кодування параметрів мовного тракту людини і джерела збудження;
- принцип кодування символної інформації (фонем);
- принцип кодування лінгвістичної інформації (слів, фраз тощо).

Дані принципи різняться в залежності від використовуваних основних властивостей мовного сигналу, освіти і сприйняття мови:

- 1) основні властивості мовного сигналу: амплітуда, частота (час), площа (як похідна амплітуди і часу);
- 2) якості створення мови:
 - зміни амплітуди;
 - розподіл мови на звуки, паузи і шуми;
 - особливості мовної та фонетичної структури;
 - короткочасна кореляція (особливості формантної структури);
 - особливості структури тону звуку (дзвінкі звуки);
 - особливості структури шуму (глухі звуки).
- 3) властивості сприйняття мови;
- 4) локальний спектральний динамічний діапазон мовного сигналу;
- 5) слухове маскування.

На сучасному етапі використовуються, в основному, підходи, які базуються на перших двох принципах кодування.

Запропонований підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови відноситься до системи електрозв'язку, а саме до систем низькошвидкісного кодування мовного сигналу, при якому здійснюється векторне квантування із прогнозом параметрів лінійного передбачення.

Метою реалізації запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови є зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування.

Зазначена мета досягається введенням в різницевий векторний квантувач параметрів лінійного передбачення передбачувача із адаптацією до основного тону мови: при класифікації кадру мовного сигналу як стаціонарного локалізованого обирають схему передбачення із адаптацією до основного тону мови. Обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення із адаптацією до основного тону мови, а вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, який дорівнює одному.

Для досягнення такого технічного результату в способі векторного квантування, що полягає в тому, що приймають вхідний вектор параметрів лінійного передбачення, класифікують кадр мовного сигналу, відповідний вхідному вектору параметрів лінійного передбачення та обчислюють вектор передбачення. Далі видаляють обчислений вектор передбачення з вхідного вектора параметрів ЛП для створення вектора похибки передбачення, масштабують вектор похибки передбачення, квантують масштабований вектор похибки передбачення.

При цьому обчислення вектора передбачення включає в себе вибір однієї з безлічі схем передбачення щодо класифікації кадру мовного сигналу і обчислення вектора передбачення відповідно до обраної схеми передбачення. Масштабування вектора похибки передбачення включає в себе вибір

щонайменше однієї з безлічі схем масштабування по відношенню до обраної схеми передбачення і масштабування вектора похибки передбачення відповідно до обраної схеми масштабування.

Множина схем передбачення включає в себе передбачення ковзаючим середнім значенням і передбачення авторегресією. При цьому якщо класифікація кадру мовного сигналу включає в себе визначення, що кадр мовного сигналу є стаціонарним вокалізованим кадром, то вибір однієї з безлічі схем передбачення включає в себе вибір передбачення авторегресією. Обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення авторегресією. Вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, і масштабування вектора похибки передбачення включає в себе масштабування вектора похибки передбачення із використанням згаданого масштабного коефіцієнта, змінено множину схем передбачення.

Множина схем передбачення включає в себе передбачення ковзаючим середнім значенням і передбачення із адаптацією до основного тону мови. При цьому якщо класифікація кадру звукового сигналу включає в себе визначення, що кадр звукового сигналу є стаціонарним вокалізованим кадром, то вибір однієї з безлічі схем передбачення включає в себе вибір передбачення із адаптацією до основного тону мови. Обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення із адаптацією до основного тону мови. Вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, рівного одному. Далі здійснюють масштабування вектора похибки передбачення.

Для опису стаціонарних вокалізованих кадрів мовного сигналу модель (1.20) може бути перетворена до наступного вигляду:

$$s(k) = b_1 \cdot s(k-1) - b_1 \cdot \Psi(s(k-1)) + G \cdot \zeta(k), \quad (2.1)$$

де $\Psi(s(k-1))$ – типове нелінійне перетворення, що визначає імпульсне збудження лінійної частини рівняння [17].

У прийнятій моделі (2.1) при $G=0$ період основного тону мови T_0 є не випадковою величиною і може бути розрахований за формулою

$$T_0 = (\ln \Delta_2 - \ln \Delta_1) / \ln b_1, \quad (2.2)$$

де Δ_1 і Δ_2 – параметри функції $\Psi(s(k-1))$.

З (2.2) залежність параметра ЛП b_1 від періоду основного тону мови T_0 може бути обчислена у такий спосіб:

$$b_1 = \exp((\ln \Delta_2 - \ln \Delta_1) / T_0). \quad (2.3)$$

Аналогічні залежності для лінійних регресій більшого порядку ($b_m, m > 1$) в аналітичному вигляді отримати досить складно, тому, з огляду на зв'язок коефіцієнтів ЛП b_i з ЛСЧ, зазначені залежності визначені експериментальним шляхом і представлені на рис. 2.1.

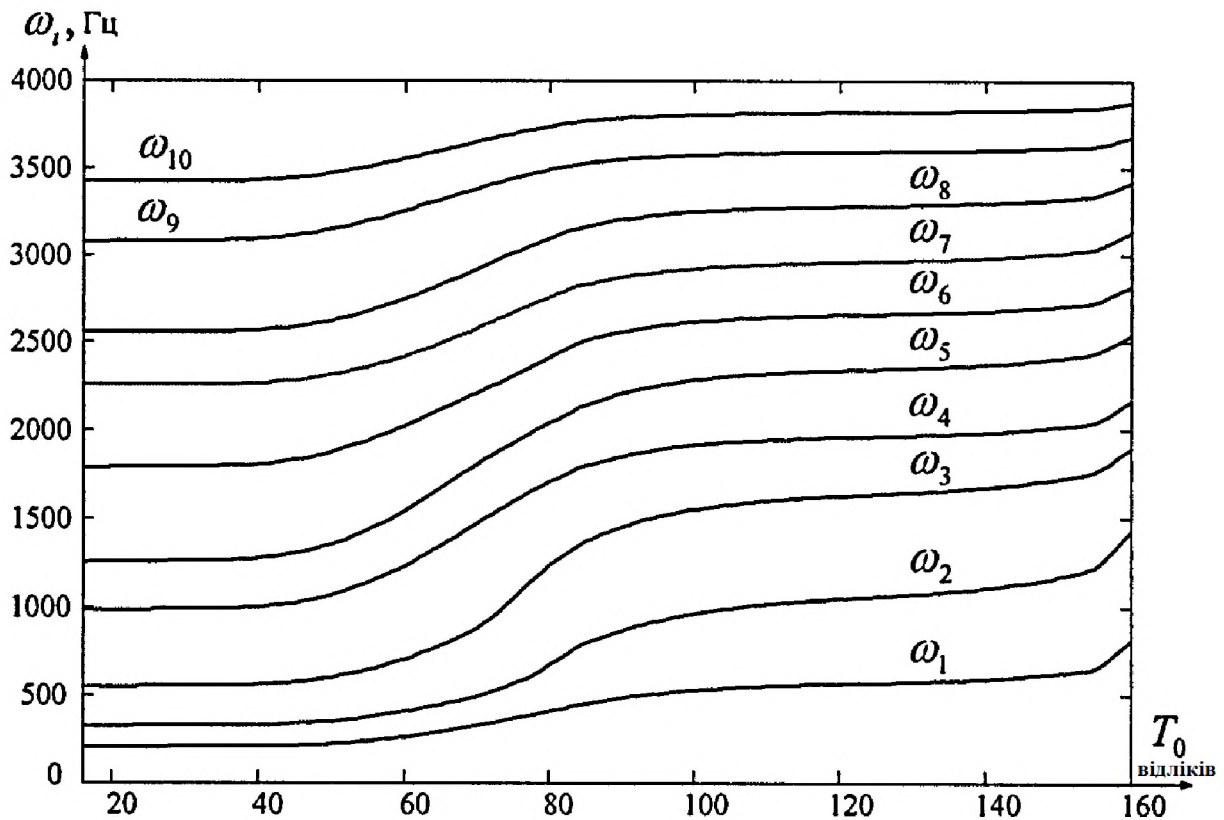


Рисунок 2.1 – Залежності ЛСЧ $\omega_i, i=1,2,\dots,M$, від періоду основного тону мови T_0 при $M=10$

Дані залежності можуть бути апроксимовані поліноміальними моделями [18] вигляду:

$$\omega_{\text{от}i} = p_1^{(i)}T_0^9 + p_2^{(i)}T_0^8 + \dots + p_{10}^{(i)} \quad (2.4)$$

де $i = \overline{1, M}$; $p_t^{(i)}$ – t -й коефіцієнт полінома, що апроксимує i -ю ЛСЧ.

Значення коефіцієнтів апроксимуючих поліномів, які використовуються для апроксимації ЛСЧ, представлені на рис. 2.2.

i	1	2	3	4	5
p_1	$2.68 \cdot 10^{-14}$	$6.76 \cdot 10^{-14}$	$-1.72 \cdot 10^{-14}$	$-4.42 \cdot 10^{-14}$	$-3.74 \cdot 10^{-14}$
p_2	$-1.47 \cdot 10^{-11}$	$-3.83 \cdot 10^{-11}$	$1.85 \cdot 10^{-11}$	$3.16 \cdot 10^{-11}$	$2.66 \cdot 10^{-11}$
p_3	$3.25 \cdot 10^{-9}$	$8.76 \cdot 10^{-9}$	$-7.22 \cdot 10^{-9}$	$-9.29 \cdot 10^{-9}$	$-7.75 \cdot 10^{-9}$
p_4	$-3.72 \cdot 10^{-7}$	$-1.03 \cdot 10^{-6}$	$1.4 \cdot 10^{-6}$	$1.45 \cdot 10^{-6}$	$1.19 \cdot 10^{-6}$
p_5	$2.36 \cdot 10^{-5}$	$6.42 \cdot 10^{-5}$	-0.000151	-0.000127	-0.0001021
p_6	-0.00087	-0.00207	0.009	0.006186	0.004748
p_7	0.021	0.031	-0.2862	-0.1543	-0.1074
p_8	-0.287	-0.0967	4.547	1.904	1.157
p_9	1.95	-1.125	-29.92	-9.632	-4.745
p_{10}	197.2	324.5	600.9	1000	1262
i	6	7	8	9	10
p_1	$-1.29 \cdot 10^{-14}$	$-2.53 \cdot 10^{-14}$	$-3.18 \cdot 10^{-14}$	$-1.94 \cdot 10^{-14}$	$-1.89 \cdot 10^{-14}$
p_2	$1.11 \cdot 10^{-11}$	$1.93 \cdot 10^{-11}$	$2.29 \cdot 10^{-11}$	$1.31 \cdot 10^{-11}$	$1.29 \cdot 10^{-11}$
p_3	$-3.73 \cdot 10^{-9}$	$-5.98 \cdot 10^{-9}$	$-6.76 \cdot 10^{-9}$	$-3.63 \cdot 10^{-9}$	$-3.59 \cdot 10^{-9}$
p_4	$-6.37 \cdot 10^{-7}$	$9.79 \cdot 10^{-7}$	$1.06 \cdot 10^{-6}$	$5.2 \cdot 10^{-7}$	$5.29 \cdot 10^{-7}$
p_5	$5.95 \cdot 10^{-5}$	$-9.06 \cdot 10^{-5}$	$-9.398 \cdot 10^{-5}$	$-4.067 \cdot 10^{-5}$	$-4.36 \cdot 10^{-5}$
p_6	0.002967	0.004688	0.004646	0.00162	0.001952
p_7	-0.072	-0.1274	-0.1193	-0.02586	-0.04317
p_8	0.8947	1.767	1.562	0.1203	0.4954
p_9	-4.753	-10.42	-8.707	0.664	-2.544
p_{10}	1798	2273	2569	3076	3422

Рисунок 2.2 – Коефіцієнти поліномів, які використовуються для апроксимації ЛСЧ

У низькошвидкісних системах кодування мовного сигналу значення періоду (частоти, нормованого періоду) основного тону підлягає передачі з кодера в декодер, тому реалізація передбачення з адаптацією до основного тону мови у вигляді виразу (2.4) дозволяє істотно зменшити динамічний діапазон ЛСЧ і, як наслідок, похибку квантування. При цьому в разі втрати мовних кадрів похибка не поширюватиметься на наступні кадри.

Для зниження кількості біт, що виділяються для кодування параметрів ЛП стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування згідно запропонованого підходу виконують наступні операції:

- 1) приймають вхідний вектор параметрів ЛП $\Omega=(\omega_1, \dots, \omega_M)$;
- 2) класифікують кадр мовного сигналу, відповідний вхідному вектору параметрів ЛП;
- 3) обчислюють вектор передбачення $\Omega_{\text{пр}}$;
- 4) видаляють обчислений вектор передбачення з вхідного вектора параметрів ЛП для створення вектора похибки передбачення E ;
- 5) масштабують вектор похибки передбачення;
- 6) квантують масштабований вектор похибки передбачення E' .

При цьому обчислення вектора передбачення включає в себе вибір однієї зі схем передбачення (прогноз ковзаючим середнім або передбачення із адаптацією до основного тону мови) по відношенню до класифікації кадру звукового сигналу і обчислення вектора передбачення відповідно до обраної схеми передбачення. Масштабування вектора похибки передбачення включає в себе вибір щонайменше однієї з безлічі схем масштабування по відношенню до обраної схеми передбачення і масштабування вектора похибки передбачення відповідно до обраної схеми масштабування.

Якщо кадр мовного сигналу класифікують як стаціонарний вокалізований, то в якості схеми передбачення вибирають передбачення із адаптацією до основного тону мови і за допомогою неї обчислюють вектор похибки передбачення $\Omega_{\text{пр}}=\Omega_{\text{от}}$, де $\Omega_{\text{от}}=(\omega_{\text{от}1}, \dots, \omega_{\text{от}M})$. В якості схеми

масштабування вибирають масштабний коефіцієнт, що дорівнює одному, що еквівалентно відсутності операції масштабування.

Рис. 2.3 ілюструє можливий варіант реалізації ВК параметрів ЛП, що використовує передбачувач із адаптацією до основного тону мови згідно із запропонованим підходом.

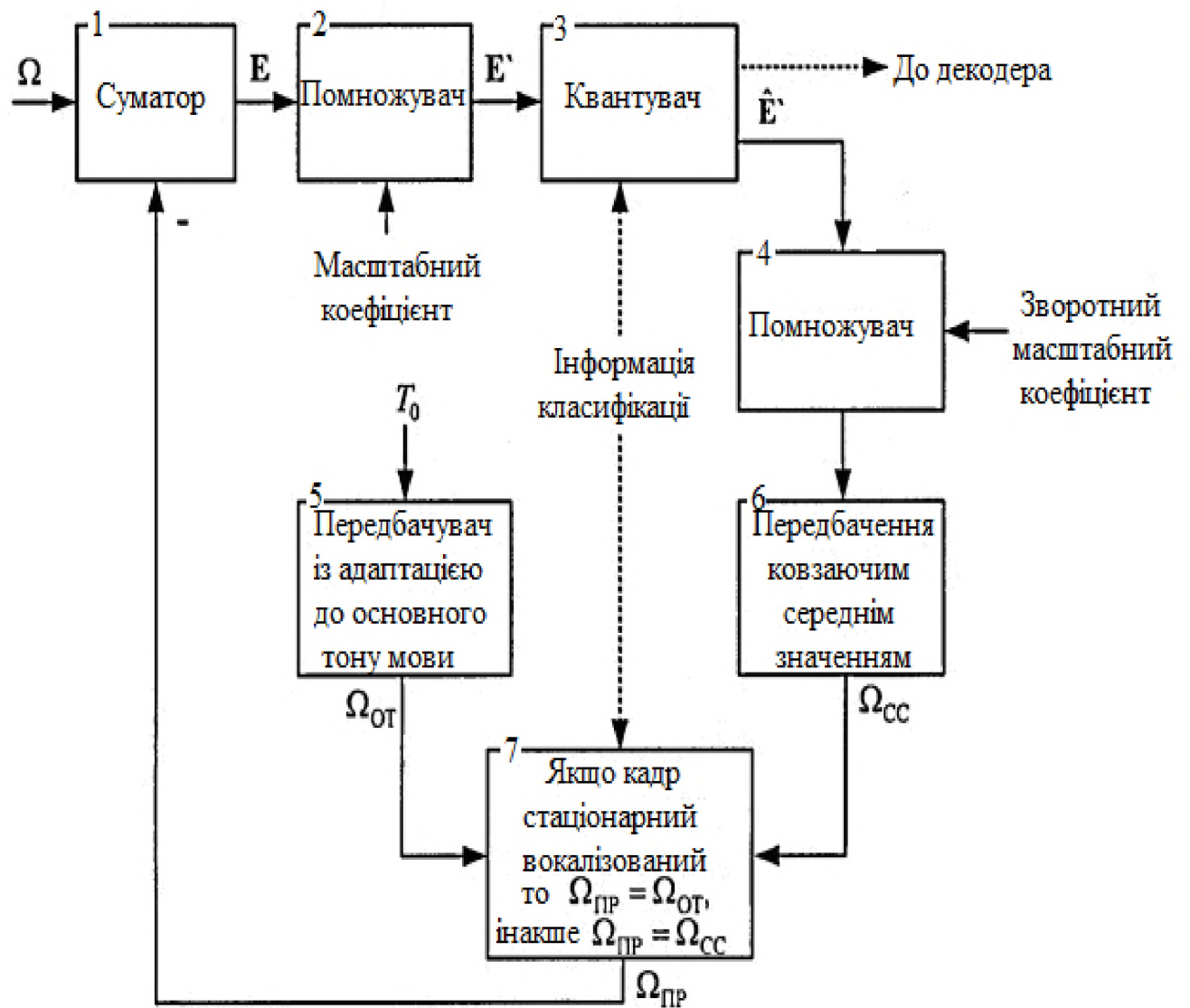


Рисунок 2.3 – Схема, що ілюструє можливий варіант реалізації векторного квантувача параметрів лінійного передбачення, згідно із запропонованим підходом

Як зображено на рис. 2.3, спочатку отримують вектор похибки передбачення E за допомогою вирахування в блоці 1 (суматор) вектора передбачення $\Omega_{\text{ПР}}$, отриманого з виходу блоку 7, з вхідного вектора параметрів ЛП Ω :

$$E = \Omega - \Omega_{\text{ПР}}. \quad (2.5)$$

Для стаціонарних вокалізованих кадрів мовного сигналу вектор передбачення $\Omega_{\text{ПР}} = \Omega_{\text{ОТ}}$ визначають в блоці 5 за поліноміальною моделлю (2.4) із коефіцієнтами, наведеними на рис. 2.2.

Для нестаціонарних мовних кадрів вектор передбачення $\Omega_{\text{ПР}} = \Omega_{\text{СС}}$ визначають за допомогою передбачення ковзаючим середнім значенням [19] в блоці 6.

Вектор похибки передбачення E масштабують в блоці 2 відповідним масштабним коефіцієнтом. При цьому для схеми передбачення із адаптацією до основного тону використовують масштабний коефіцієнт, що дорівнює 1, а для схеми передбачення ковзаючим середнім значенням – масштабний коефіцієнт більше 1.

Квантоване значення \hat{E}' масштабованого вектора похибки передбачення E' отримане в блоці 3, після зворотного масштабування в блоці 4 використовують в блоці 6 передбачення ковзаючим середнім значенням, а його індекс в КК передають в декодер.

2.2 Оцінка ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови

Оцінка ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови проводилась в середовищі Matlab / Simulink за допомогою стандартних і розроблених програм.

Для перевірки працездатності розглянутого ВК і оцінки якості квантування параметрів ЛП згідно запропонованого підходу було здійснено імітаційне моделювання.

Якість квантування параметрів ЛП оцінювалася з використанням спектрального спотворення [20], що визначається наступним чином:

$$sd = \sqrt{\frac{1}{\pi} \int_0^{\pi} \left(10 \lg \frac{P(\omega)}{\hat{P}(\omega)} \right)^2 d\omega}, \quad (2.5)$$

де P_j та \hat{P}_j – спектри потужності мовного кадру, розраховані за вхідними і квантованими ЛСЧ відповідно.

Середнє спектральне спотворення оцінювалось як

$$SD = \frac{1}{J} \sum_{i=1}^J sd(i), \quad (2.6)$$

де J – число вимірювань (кількість аналізованих кадрів мовного сигналу).

В [21] встановлено такі умови, що призводять до відсутності чутних спотворень через спектральні невідповідності:

- середнє значення спектрального спотворення SD не перевищує 1 дБ;
- кількість кадрів, на яких sd приймає значення від 2 до 4 дБ, не перевищує 2%;
- відсутність сегментів зі значенням sd більше 4 дБ.

Навчання кодової комбінації масштабованого вектора похибки передбачення E' здійснювалось за допомогою узагальненого алгоритму Ллойда [22-24]. Для навчання і тестування КК використовувались записи мови різних дикторів тривалістю близько 125 і 15 хвилин відповідно. При цьому мова в навчальній і тестовій вибірках належала різним дикторам, а загальне число векторів ЛСЧ в них склало 375873 і 45318 відповідно (співвідношення тривалостей вибірок при цьому приблизно дорівнює 8:1). Значення періоду основного тону мови передбачалось відомим.

Порівняльна оцінка якості квантування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу для різного числа біт згідно підходу-прототипу і запропонованого підходу до векторного

квантування параметрів лінійного передбачення із адаптацією до основного тону мови представлена на рис. 2.4.

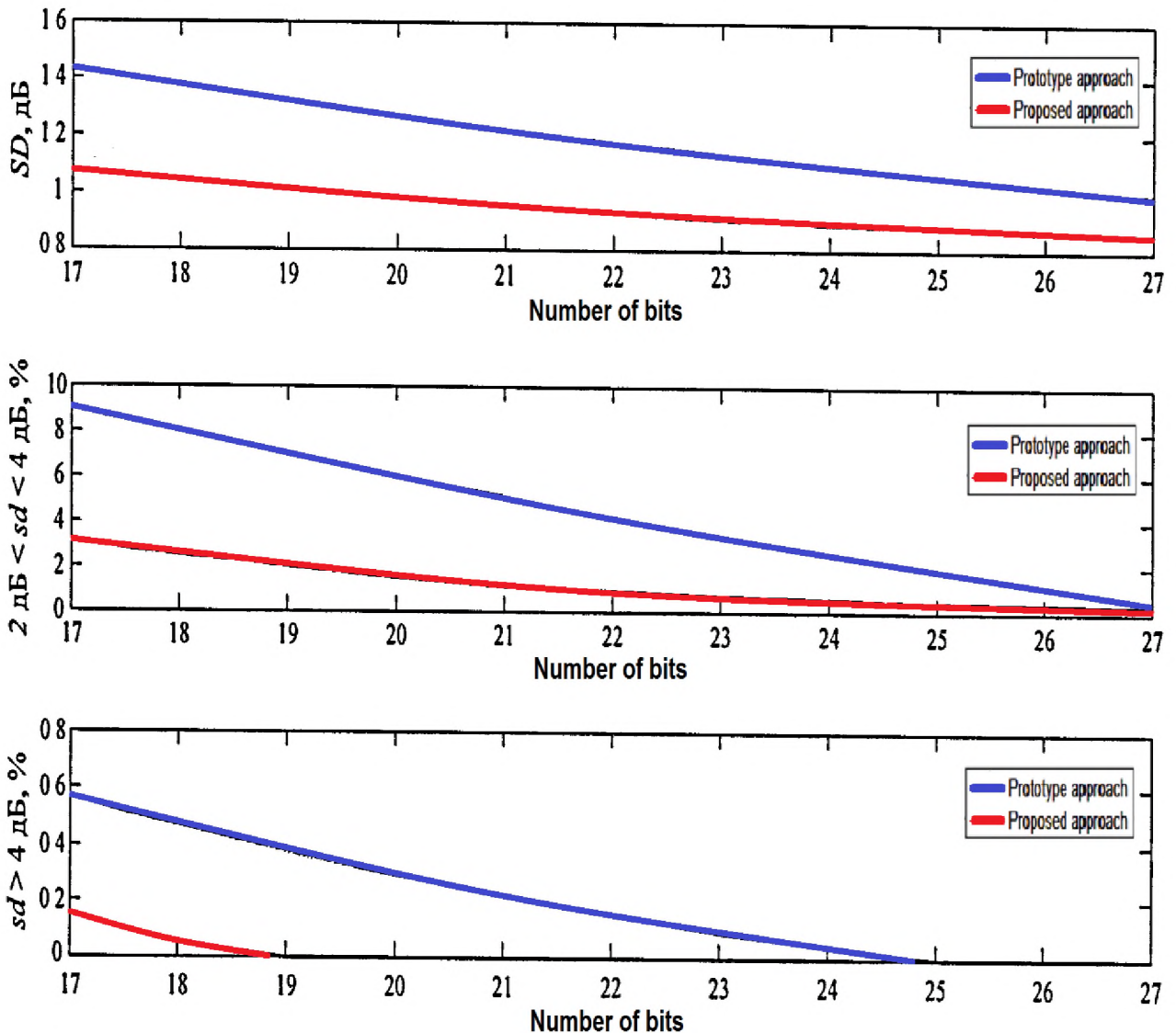


Рисунок 2.4 – Порівняльна оцінка якості квантування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу для різного числа біт згідно прототипу і запропонованого підходу із адаптацією до основного тону мови

В результаті оцінки ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови (згідно рис. 2.4) встановлено, що ВК із адаптацією до основного тону мови відповідно до запропонованого підходу дозволяє знизити кількість біт,

необхідних для кодування параметрів ЛП стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування у порівнянні із відомим підходом-прототипом, який використовує передбачення авторегресією.

2.3 Висновки

Запропонований підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови відноситься до системи електрозв'язку, а саме до систем низькошвидкісного кодування мовного сигналу, при якому здійснюється векторне квантування із прогнозом параметрів лінійного передбачення.

Метою реалізації запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови є зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування.

Зазначена мета досягається введенням в різницевий векторний квантувач параметрів лінійного передбачення передбачувача із адаптацією до основного тону мови: при класифікації кадру мовного сигналу як стаціонарного локалізованого обирають схему передбачення із адаптацією до основного тону мови. Обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення із адаптацією до основного тону мови, а вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, який дорівнює одному.

Для оцінки ефективності запропонованого підходу із адаптацією до основного тону мови (перевірки працездатності розглянутого ВК і оцінки якості квантування параметрів ЛП) було здійснено імітаційне моделювання в середовищі Matlab / Simulink за допомогою стандартних і розроблених програм. При цьому, якість квантування параметрів ЛП оцінювалося з використанням спектрального спотворення.

Встановлено, що ВК із адаптацією до основного тону мови відповідно до запропонованого підходу дозволяє знизити кількість біт, необхідних для кодування параметрів ЛП стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування у порівнянні із відомим підходом-прототипом, який використовує передбачення авторегресією.

3 ЕКОНОМІЧНА ЧАСТИНА

3.1 Визначення трудомісткості векторного квантування параметрів лінійного передбачення

Трудомісткість – це кількість робочого часу, що витрачається на створення одиниці продукції. Трудомісткість обернено пропорційна показнику продуктивності праці (кількості продукції, що виробляється за одиницю робочого часу). Поняття трудомісткості тісно пов'язане з поняттям капіталоємності (кількості капіталу, що витрачається на виробництво одиниці продукції).

Трудомісткість векторного квантування параметрів лінійного передбачення визначається тривалістю кожної робочої операції (таблиця 3.1), починаючи зі збору та аналізу необхідної інформації і закінчуючи оформленням документації (за умови роботи одного чоловіка):

$$t = t_1 + t_2 + t_3 + t_4 + t_5 + t_6, \text{ годин} \quad (3.1)$$

Оцінка витрат праці на збір і аналіз інформації залежить від конкретних умов і визначається на основі експертних оцінок. Зважаючи на той факт, що дослідження, пов'язані з обробкою складних сигналів в телекомунікаційних системах, охоплюють великий пласт інформації, приймаємо:

Таблиця 3.1 – Тривалість кожної робочої операції векторного квантування параметрів лінійного передбачення

№	Назва робочої операції	Тривалість здійснення операцій
1.	Збір і аналіз інформації	$t_1=40$ годин
2.	Постановка задачі	$t_2=2$ години
3.	Складання технічного завдання	$t_3=7$ годин
4.	Підхід до векторного квантування параметрів лінійного передбачення із адаптацією до	$t_4=17$ годин

	основного тону мови	
5.	Оцінка ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення	$t_5=21$ годин
6.	Підготовку документації по завданню	$T_6=22$ годин

Отже, трудомісткість при дослідженні і розробці підходу до векторного квантування параметрів лінійного передбачення складає 102 години:

$$t = 40 + 2 + 17 + 7 + 21 + 22 = 102 \text{ годин.}$$

3.2 Визначення середньої заробітної плати спеціаліста в галузі телекомунікацій

До обов'язків інженера з телекомунікацій входить забезпечення доступу користувачам в мережу інтернет, прокладання оптико-волоконних мереж, надання послуги IP-телефонії. Також він відповідає за налагодження, експлуатацію, моніторинг мережі, її розвиток, забезпечує якісну і надійну роботу систем, встановлює на підприємствах системи відеоспостереження, охоронно-пожежної сигналізації, звукового оповіщення та ін.

Середня зарплата в категорії «Телекомунікації та зв'язок» в Україні - 15000 грн. Це медіана заробітних плат за даними з 10860 вакансій, розміщених на Work.ua в цій категорії за останні 3 місяці [25].

Розрахуємо середню заробітну плату за одну годину роботи, з урахуванням 8-ми годинного робочого графіку на добу і 5-ти денного робочого тижня:

Кількість робочих годин в місьць (підррахунок за місяць квітень, 30 днів в місяці):

$$t = 8 \text{ годин} \cdot 21 \text{ день} = 168 \text{ год/місяць}$$

Середня заробітна плата за одну годину роботи становить:

$$ЗП_{\text{сер/год}} = 15000 / 168 = 89,29 \text{ грн/годину}$$

3.3 Розрахунок витрат на заробітну плату при дослідженні векторного квантування параметрів лінійного передбачення

Заробітна плата (оплата праці працівника) – винагорода за працю залежно від кваліфікації працівника, складності, кількості, якості та умов виконуваної роботи, а також компенсаційні і стимулюючі виплати. Заробітна плата - грошова компенсація, яку працівник отримує в обмін на свою працю.

Заробітна плата виконавця враховує основну і додаткову заробітну плату, а також відрахування на соціальні потреби (єдиний соціальний внесок) і визначається за формулою:

$$Z_{зп} = t \cdot Z_{іт}, \text{ грн}, \quad (3.4)$$

де t – загальна тривалість при дослідженні векторного квантування параметрів лінійного передбачення, годин;

$Z_{іт}$ – середньогодинна заробітна плата інженера в галузі телекомунікацій, грн/годину.

$$Z_{зп} = 102 \cdot 89,29 = 9107,14 \text{ грн.}$$

3.4 Розрахунок витрат на необхідне програмне та апаратне забезпечення

У процесі реалізації програмних проектів операційна система використовується на всіх етапах проектування програм, починаючи від формулювання алгоритмів та структурних схем і закінчуючи відладкою та виконанням програм.

До складу початкового програмного забезпечення (ПО) даного комп'ютера входить операційна система (ОС), а також стандартний пакет прикладних програм і драйверів.

Вартість необхідного програмного та апаратного забезпечення представлена в табл. 3.2.

Таблиця 3.2 – Вартість необхідного програмного та апаратного забезпечення

Найменування	Вартість, грн
Ноутбук Ноутбук Asus Laptop X509JB-EJ063 Grey	15699,00
Миша A4Tech G7-360N	390,00
Операційна система Microsoft Windows 10, 32-bit	3699,00
Разом:	19788

3.5 Висновки

В економічному розділі розрахований час, який необхідний для розробки підходу до векторного квантування параметрів лінійного передбачення, він становить 102 години. Заробітна плата працівника за час виконання дослідження – 9107,14 грн. При чому вартість програмного та апаратного забезпечення складає 19788,00 грн.

ВИСНОВКИ

1. В результаті аналізу основних положень формування і обробки мовних сигналів встановлено, що у зв'язку з активним розвитком техніки зв'язку, рішення задачі кодування мовної інформації має велике значення. Встановлено, що кодування з лінійним передбаченням є одним з найпотужніших методів аналізу мови, і одним з найкорисніших для кодування хорошої якості мови при низькій швидкості передачі бітів. До того ж LPC забезпечує надзвичайно точні оцінки параметрів мови.

2. В результаті аналізу існуючих підходів до векторного квантування параметрів лінійного передбачення встановлено їх недоліки. Недоліком відомого підходу до векторного квантування параметрів ЛП CS-ACELP [14] є велике значення похибки квантування, що представляє собою різницю між квантованим і початковим вектором параметрів ЛП. Слід зауважити, що зазначений недолік обумовлений малим числом біт, що відводиться на квантування ЛСЧ в низькошвидкісних системах кодування мовного сигналу.

Недоліками відомого підходу до перетворення мовного сигналу методом лінійного передбачення з адаптивним розподілом інформаційних ресурсів [15] є значна похибка квантування параметрів ЛП, збільшення числа переданих біт, а також артефакти звучання синтезованої мови, викликані перемиканням ВК для кожного класу кадрів мовного сигналу.

Недоліками відомого підходу для векторного квантування з надійним прогнозом параметрів лінійного передбачення в кодуванні мови зі змінною бітовою швидкістю [16] (який було обрано як прототип) є наступні:

1) передбачення авторегресією, що застосовується для стаціонарних вокалізованих кадрів, не є надійним в умовах втрати кадрів – в разі втрати мовних кадрів похибка поширюється на наступні кадри;

2) для отримання мінімальної похибки квантування параметрів ЛП стаціонарних вокалізованих кадрів потрібно витратити значне число біт.

3. Запропоновано підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови, який забезпечує зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування. Це досягається введенням в різницевий векторний квантувач параметрів лінійного передбачення передбачувача із адаптацією до основного тону мови: при класифікації кадру мовного сигналу як стаціонарного локалізованого обирають схему передбачення із адаптацією до основного тону мови. Обчислення вектора передбачення включає в себе обчислення вектора похибки передбачення за допомогою передбачення із адаптацією до основного тону мови, а вибір однієї з безлічі схем масштабування включає в себе вибір масштабного коефіцієнта, який дорівнює одному.

4. В результаті оцінки ефективності запропонованого підходу до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови (згідно рис. 2.4) встановлено, що ВК із адаптацією до основного тону мови відповідно до запропонованого підходу дозволяє знизити кількість біт, необхідних для кодування параметрів ЛП стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування у порівнянні із відомим підходом-прототипом, який використовує передбачення авторегресією.

ПЕРЕЛІК ПОСИЛАНЬ

1. Конахович Г.Ф. Системи радіозв'язку. Навчальний посібник. / Г.Ф. Конахович. – К.: НАУ, 2004 – 311с.
2. Игнатов В.О. Принципы построения эксплуатации систем связи воздушных судов: Учебное пособие / В.О. Игнатов, Г.Ф. Конахович, В.В. Уланський, Ю.М. Омелько. – К.: Книга, 1988. – 172 с.
3. Шелухин О.И. Цифровая обработка и передача речи / О.И. Шелухин, Н.Ф. Лукьянцев. – М.: Радио и связь, 2000. – 456 с.
4. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М.: Радио и связь, 1981. – 206 с.
5. Бабков В.Ю. Системы связи с кодовым разделением каналов / В.Ю. Бабков, М.А. Вознюк. – СПб.:, 1999. – 120 с.
6. Варакин Л.Е. Системы связи с шумоподобными сигналами / Л.Е. Варакин. – М.: Радио и связь, 1985. – 384 с.
7. Феер К. Беспроводная цифровая связь. Методы модуляции и расширение спектра. Пер. с англ. // Под. ред. В.И. Журавлева. – М.: Радио и связь, 2000. – 520 с.
8. Гуров В.С. и др. Передача дискретной информации и телеграфия. – М.: Связь, 1984. – 526 с.
9. Сапожков М.А. Вокодерная связь / М.А. Сапожков, В.Г. Михайлов. – М.: Радио и связь, 1983. – 248 с.
10. Анисимов В.А. Системы авиационной радиосвязи / В.А. Анисимов. – М.: Транспорт, 1981. – 239 с.
11. Андрусак А.І. Мережа авіаційного електровз'язку / А.І. Андрусак, В.С. Дем'янчук, Ю.М. Юр'єв. – К.: НАУ, 2001. – 448 с.
12. Голд Б. Цифровая обработка сигналов : пер. с англ. / Б. Голд, Ч. Рэйдер; под ред. А.М. Трахмана. – М. : Сов. Радио, 1973. – 368 с.
13. Быков С.Ф. Цифровая телефония: учеб. пособие для вузов / С.Ф. Быков, В.И. Журавлев, И.А. Шалимов. – М.: Радио и связь, 2003. – 144 с.

14. Рекомендації ITU-T G.729 «Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)». – Женева. – 1996.
15. Патент РФ 2248619. Способ и устройство преобразования речевого сигнала методом линейного предсказания с адаптивным распределением информационных ресурсов / А.А. Рыболовлев, Г.В. Богачев, В.Г. Трубицын, И.А. Азаров. – Заявл. 27.08.2004, опубл. 20.03.2005.
16. Патент РФ 2326450. Способ и устройство для векторного квантования с надежным предсказанием параметров линейного предсказания в кодировании речи с переменной битовой скоростью / М. Желинек. – Заявл. 24.12.2006, опубл. 10.06.2008.
17. Прохоров Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов / Ю.Н. Прохоров. – М.: Радио и связь, 1984. – 240 с.: ил.
18. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников: учеб. пособие / А.И. Кобзарь. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
19. Koestoer N.P. Robust linear prediction analysis for low bit-rate speech coding: DPh dissertation / N.P. Koestoer. – Brisbane: Griffith University, 2002. – 132 p.
20. Павловец А.Н. Квантование огибающей спектра в вокодер, основанное на декомпозиции речевого сигнала на периодическую и аperiodическую составляющие / А.Н. Павловец, А.А. Петровский // Цифровая обработка сигналов. – 2005. – №3. – С.13-21.
21. Palival K.K. Efficient vector quantization of LPC parameters at 24 bits/frame / K.K. Palival, B.S. Atal // IEEE Trans. on Acoustics, Speech and Signal Processing. – 1993. – Vol.1. – №1. – P.3-14.
22. Макхоул Д. Векторное квантование при кодировании речи / Д. Макхоул, С. Рукос, Г. Гиш // ТИИЭР. – 1985. – Т.73. – №11. – С.19-61.
23. Ахмад, Х.М. Введение в цифровую обработку речевых сигналов : учеб. пособие / Х.М. Ахмад, В.Ф. Жирков. – Владимир : Изд-во Владим. гос. ун-та, 2007. – 192 с. – ISBN 5-89368-751-5.

24. Методичні рекомендації до виконання дипломних робіт (проектів) бакалаврів та магістрів спеціальностей 125 Кібербезпека, 172 Телекомунікації та радіотехніка / Упоряд.: О.Ю. Гусєв, О.В. Герасіна, О.М. Алексєєв, О.В. Кручінін. – Дніпро: НГУ, 2018. – 50 с.

25. Средняя зарплата по категории «Телекоммуникации и связь» в Украине. <https://www.work.ua/ru/salary-telecommunications/>

ДОДАТОК А. Відомість матеріалів кваліфікаційної роботи

№	Формат	Найменування	Кількість листів	Примітки
<i>Документація</i>				
1	A4	Реферат	3	
2	A4	Список умовних скорочень	1	
3	A4	Зміст	2	
4	A4	Вступ	2	
5	A4	Стан питання. Постановка задачі	37	
6	A4	Спеціальна частина	12	
7	A4	Економічний розділ	4	
8	A4	Висновки	2	
9	A4	Перелік посилань	3	
10	A4	Додаток А	1	
11	A4	Додаток Б	1	
12	A4	Додаток В	1	
13	A4	Додаток Г	1	

ДОДАТОК Б. Перелік документів на оптичному носії

1 Презентація Синій.ppt

2 Диплом Синій.doc

ДОДАТОК Г. Відгук керівника кваліфікаційної роботи

В І Д Г У К

**на кваліфікаційну роботу студента групи 172-18ск-1 Синього Д.С.
на тему: «Векторне квантування параметрів лінійного передбачення
мовного сигналу»**

Пояснювальна записка складається зі вступу, трьох розділів і висновків, розташованих на 72 сторінках.

Мета роботи є актуальною, оскільки вона спрямована на зниження кількості біт, що виділяються для кодування параметрів лінійного передбачення стаціонарних вокалізованих кадрів мовного сигналу, при мінімальній похибці квантування.

При виконанні роботи автор продемонстрував добрий рівень теоретичних знань і практичних навичок. На основі аналізу основних положень формування і обробки мовних сигналів, а також існуючих підходів до векторного квантування параметрів лінійного передбачення в ній сформульовані задачі, вирішенню яких присвячений спеціальний розділ. У ньому було запропоновано підхід до векторного квантування параметрів лінійного передбачення із адаптацією до основного тону мови та оцінено його ефективність.

Практична цінність роботи полягає в тому, що запропонований підхід може бути використаний для скорочення обсягу цифрового мовного сигналу.

До недоліків роботи слід віднести недостатню проробку окремих питань.

Рівень запозичень у кваліфікаційній роботі відповідає вимогам «Положення про систему виявлення та запобігання плагіату».

В цілому робота задовольняє усім вимогам, а її автор Синій Д.С. заслуговує на оцінку «
» та присвоєння кваліфікації «Бакалавр з телекомунікацій та радіотехніки» за спеціальністю 172 Телекомунікації та радіотехніка.

**Керівник роботи,
к.т.н., доцент**

О.В. Герасіна