МІНІСТЕРСТВО ОСВІТИ І НАУКИ
ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ЗВ'ЯЗКУ ім. О.С. ПОПОВА

Є. М. Рудий

ТЕХНОЛОГІЇ ПЕРЕДАЧІ ДИСКРЕТНИХ ПОВІДОМЛЕНЬ

Модуль 1: Стиснення телефонних сигналів

Навчальний посібник для освітньо-професійної підготовки бакалаврів
за напрямом вищої освіти 6.050903 – Телекомунікації

Одеса
2013

УДК 621.391.037.372:534.863
ББК 32.811.3 План НМВ 2012/2013 уч. р.
Р-83

Рецензенти:
П. Ю. Баранов, д.т.н., професор (Одеський національний політехнічний університет).
М. М. Климаш, д.т.н., професор (Національний університет «Львівська політехніка»).

Рудий Є. М. **Технології передачі дискретних повідомлень.** Модуль 1: **Стиснення телефонних сигналів:** Навчальний посібник. – Одеса: ОНАЗ ім. О.С. Попова, 2013. – 102 с.

Відповідальний редактор М. В. Захарченко

Розглянуті методи стиснення телефоних сигналів, які пристосовуються до виду сигналу та дозволяють стиснути сигнал за рахунок інформаційної надмірності. Описані принципи побудови систем стиснення сигналів, які використовують диференційні, спектрально-смугові, формантні, ортогональні, кореляційні та мовоелементні методи.

Призначений для студентів, що навчаються за напрямом 6.050903 «Телекомунікації», а також може бути корисним інженерно-технічним працівникам.

ЗАТВРЕДЖЕНО
методичною радою академії
Протокол № 3/14
від 9 квітня 2013 р.

Навчальний посібник
розглянуто та ухвалено
на засіданні кафедри
інформаційної безбеки та передачі даних.
Протокол № 7 від 31 січня 2013 року.

ВСТУП

Для підвищення ефективності використання каналів зв′язку застосовують різні методи та засоби. Серед них важливу роль відіграють методи стиснення сигналів, яким присвячені перші два модулі дисципліни «Технології передачі дискретних повідомлень». Стиснення сигналів забезпечується за рахунок інформаційної надмірності. Це призводить до зменшення цифрового потоку і затрат на оренду каналів зв'язку.

Значна увага в даному посібнику приділяється системам стиснення мовних сигналів, які називають вокодерами (*voice* – голос, *coder* – кодувальник). Застосування вокодерів дозволяє підвищити також завадостійкість телефонного зв′язку і забезпечити таємність розмов. Широкому впровадженню вокодерів у техніку зв′язку сприяли цифрові методи обробки сигналів.

Спотворення при стисненні звукових сигналів можуть бути помітними або непомітними при слуховому сприйнятті. І саме тому, що існують спотворення, які не помічаються вухом людини, можливе стиснення не тільки сигналів телефонного зв′язку, але й сигналів звукового мовлення. Не можна дозволяти того, щоб під час передачі програм звукового мовлення були відчутні спотворення, викликані системою стиснення.

Трохи інакше підходять до спотворень при передачі сигналів телефонного зв′язку. Основним параметром якості при передачі мовних сигналів є розбірливість, бо якщо канал зв′язку не забезпечує повного розуміння мови, то ніякі інші переваги такого каналу зв′язку не мають вже значення – він не придатний до експлуатації. Мова людини має значну надмірність. По голосу можна визначити особу абонента, його настрій, стать, вік, стан здоров′я тощо. Якщо з мовного сигналу виділити тільки параметри, що визначають розбірливість мови, то за сучасної техніки зв′язку в одному телефонному каналі можна розмістити близько 160 вокодерних каналів. Така техніка використовується в комерційному та військовому зв′язку, при цьому забезпечується більша розбірливість ніж в телефонних мережах зі звичайними апаратами. Якщо дати можливість абонентам пізнавати один одного по голосу, то на даному етапі розвитку техніки стиснення замість одного телефонного каналу передають близько 100 вокодерних каналів.

При розробці систем стиснення звукових сигналів треба враховувати властивості вуха людини та характеристики акустичних сигналів, тому вони коротко описуються в даному посібнику.

Автор щиро вдячний доктору технічних наук, професору М.В. Захарченко, поради та зауваження якого враховані при написанні посібника.

1. СТИСНЕННЯ ДАНИХ

Стискати обсяг інформації, яка передається або запам′ятовується, можна тільки в тому випадку, коли в інформаційному сигналі існує надмірність. Причини виникнення надмірності такі: 1) інформація, яка передається по каналу зв′язку, не в повному обсязі враховує психофізичні якості абонента, якому необхідно менше відомостей для сприйняття інформації; 2) неоптимальні методи передачі інформації приводять до ненавмисного збільшення її обсягу; 3) стиснення даних організоване таким чином, що не реалізуються оптимальні методи обробки сигналів; 4) неповні дані про характер переданої інформації приводять до того, що в повідомленні виникає природна надмірність; 5) при резервуванні інформаційних каналів, а також при використанні завадостійкого кодування вноситься навмисна надмірність, яка повністю контролюється розробником апаратури [1].

Першою ознакою класифікації методів стиснення даних можуть бути причини виникнення надмірності в переданій інформації.

Методи стиснення можна також класифікувати за типом повідомлень (аналогові, аналого-дискретні, цифрові).

Друга ознака класифікації – кількість параметрів, що можуть одночасно оброблятись системою стиснення. Між такими параметрами можливий детермінований або випадковий зв′язок, а при спільній обробці параметрів можна врахувати ці зв′язки, що приведе до додаткового виграшу при стисненні.

Третя ознака класифікації – орієнтація методів стиснення на певний клас об′єктів, які враховують специфіку цих об′єктів (стиснення телевізійних сигналів, текстової інформації, сигналів звукового мовлення, мовних сигналів).

Четверта ознака класифікації – кількість кодових слів, що одночасно обробляються системою стиснення. Коли система стиснення опрацьовує тільки одне кодове слово, то технічна реалізація відносно проста, але розмір стиснення сигналів малий. При використанні систем стиснення з одночасним опрацюванням груп слів отримають більше стиснення. Ці методи складні при технічній реалізації, але стиснення забезпечується за невеликих спотворень сигналів.

Системи стиснення можна також класифікувати за можливістю відновлення вихідного повідомлення з заданою похибкою, тобто розрізняють оборотні та необоротні методи стиснення. Оборотні методи дозволяють відновлювати параметри повідомлень з заданою ймовірністю. Необоротні методи не дозволяють відновлювати параметри вихідного повідомлення [2].

Можлива й інша класифікації методів стиснення даних, за якої враховуються ознаки конкретних схем стиснення.

Дослідження систем стиснення можна проводити експериментально й аналітично. Складність технічної реалізації алгоритмів стиснення сигналів призводить до того, що експериментальні дослідження поступово витискуються аналітичними. Часто використовують фізико–математичне моделювання систем стиснення за допомогою ЕОМ. Таке моделювання дозволяє при дослі-

дженнях використовувати реальні інформаційні сигнали та різко зменшити строки проведення досліджень. Якщо алгоритм стиснення добре опрацьований при моделюванні на ЕОМ, то можлива технічна реалізація спеціалізованих процесорів систем стиснення. В цьому випадку немає необхідності виготовлення та настроювання громіздких макетів апаратури. Саме з використанням фізико–математичного моделювання в Одеській національній академії зв′язку ім. О. С. Попова виконувались дослідження з стиснення сигналів звукового мовлення.

Контрольні питання

1. В якому випадку можна стискати інформаційний сигнал?
2. Які причини виникнення надмірності інформаційних сигналів?
3. Які ознаки методів стиснення даних?
4. Як проводять дослідження систем стиснення?

2. ОСНОВНІ ВЛАСТИВОСТІ СЛУХУ ТА ХАРАКТЕРИСТИКИ ЗВУКОВИХ СИГНАЛІВ

Апаратуру, призначену для передачі звукових сигналів, слід проектувати з урахуванням властивостей вуха людини. Необхідно знати, які складові сигналів є інформативними, які спотворення сигналів сприймаються слухом, як це пов′язане з якістю передачі музичних творів, як та чи інша обробка сигналів сприяє розбірливості мови. Знання характеристик слуху дозволяє сформулювати вимоги і до систем стиснення сигналів.

У вусі людини є мембрана, в якій розміщені волокна різної довжини. Волокна закінчуються тонкими волосками, які являють собою чутливі елементи волоскових клітин.

Згідно з резонансною теорією слуху, запропонованою визначним фізиком і лікарем Г. Гельмгольцем, волокна мембран є набором зі значної кількості резонаторів. Кожний із цих резонаторів відгукується (резонує) при надходженні коливань визначеної частоти. Довгі волокна резонують на низьких частотах, а короткі волокна – на високих частотах. Через те, що волокна пов′язані між собою, окремі резонансні коливання волокон практично неможливі. Волокна зв′язані зі слуховим нервом, по якому інформація про звукові коливання подається до слухового центру мозку.

Кількість нервових волокон складає близько 3000, тому кількість подразнень, на які реагує слух людини, дуже значна, дозволяючи людині розрізняти тонкі особливості в звукових сигналах за їх інтенсивністю, частотою коливань та спектральним складом.

Основні властивості слуху були отримані в результаті проведення статистичних експериментальних досліджень, які проводились у багатьох країнах світу. Був визначений поріг чутності за різних частот звукових сигналів. Порогом чутності називають значення інтенсивності звукового синусоїдального сигналу, яке викликає у повній тиші відчуття звуку. Звуковий сигнал, який дорівнює порогові чутності, відчувають 50% людей, що брали участь у лабораторних дослідженнях, віком від 18 до 23 років. Отже, у 50% досліджених поріг більш низький (слух більш гострий), а у 50% – поріг більш високий [3-4].

На частоті 1000 Гц поріг чутності дорівнює 10^{-12} Вт / м^2. Звукові сигнали у межах від 2000 до 4000 Гц відчутні за інтенсивності звуку менш ніж 10^{-12} Вт / м^2. Далі, при підвищенні частоти, поріг чутності збільшується, а при частоті 20000 Гц він у 10^6 разів вищий, ніж при частоті 1000 Гц. І якби не збільшувався за інтенсивністю звук з частотою вище 20000 Гц, він не сприймається слухом.

Також не сприймаються звуки з частотою нижче 16…20 Гц. На частоті 50 Гц поріг чутності у $5 \cdot 10^5$ разів більший, ніж на частоті 1000 Гц.

Виходячи з величини порога чутності на частоті 1000 Гц, визначають рівень L звукового сигналу

$$L = 10 \lg \frac{I}{I_0}, \qquad (2.1)$$

де $I_0 = 10^{-12}$ Вт / м^2, I – інтенсивність або сила звуку.

Інтенсивність звукового сигналу це середня кількість енергії, що проходить за одиницю часу через одиничну площадку, розміщену перпендикулярно до напрямку поширення звукових хвиль. Поріг чутності на частоті 1000 Гц відповідає рівню звукового сигналу 0 дБ.

Люди віком від 18 до 23 років сприймають звукові сигнали з частотами 16…20000 Гц. Люди старшого віку чують гірше як низькі, так і високі частоти. У людей які знаходяться під постійним впливом гучних звуків, поріг чутності підвищується (людина втрачає слух). Тому сільські мешканці мають менший поріг чутності, ніж міські. Поріг чутності різко підвищується у молодих людей, які захоплюються поп-музикою, для неї характерні великі рівні звучання. В даному випадку погіршення слуху є захисною реакцією організму на дуже потужний подразник, яким є гучні звукові сигнали.

У літературі наведені різні абсолютні значення і частотні залежності порогу чутності. При бінауральному прослуховуванні поріг чутності менше, ніж при моноуральному [5].

Якщо збільшити сигнал даної частоти, то сприйняття сигналу збільшується і за достатньо значною інтенсивності звукового сигналу настає відчуття болю у вухах. Таку величину сигналу називають ***порогом больового відчуття***. На частоті 1000 Гц поріг больового відчуття настає при рівнях 120…130 дБ.

Слухове відчуття при збільшенні звуку зростає не плавно, а стрибками. Такі стрибки називають порогами розрізнення інтенсивності. В області середніх частот кількість стрибків близько 250. На низьких та середніх частотах кількість порогів розрізнення менша, а в середньому по частотному диапазоні складає близько 150.

Дискретність слуху спостерігається не тільки за амплітудою звукових сигналів, але й за частотою. В усьому частотному діапазоні людина сприймає не більше 250 градацій частоти, причому кількість градацій залежить від інтенсивності звукових сигналів. В області середніх частот людина може розпізнати зміну частоти до 0,3%, якщо частота змінюється плавно. Якщо дискретно змінювати частоти сигналів, то людина розпізнає частоти гірше. Наприклад, найкращі музиканти з абсолютним слухом не виявляють різниці в звучанні, якщо фільми, зняті для кіно, демонструються на телебаченні та навпаки. В телебаченні зміна кадрів проводиться зі швидкістю 25 кадрів за секунду, а у кіно – зі швидкістю 24 кадри за секунду. Розходження частот звукових сигналів в такому випадку досягає 4% [5].

Дискретність слухового сприйняття обумовлена кінцевим числом (22000) нервових закінчень – волосків, які підходять до волокон, розміщених у мембрані. Оскільки людина розпізнає звукові сигнали одночасно за амплітудою та частотою, то кількість градацій (стрибків) за частотою та амплітудою буде $\sqrt{22000} \approx 150$. Це твердження доведене експериментально.

У 1846 році Вебером було установлене загальне психофізіологічне співвідношення, яке стверджує, що мінімально помітний приріст подразника складає близько 10% від початкової інтенсивності подразника. Це пов′язане з властивостями нервової системи і справедливе для звукових, світлових та інших подразників, тому називається ***загальним фізіологічним законом***. Математичне формулювання співвідношення Вебера виконане Фехнером у 1860 році і називається ***психофізіологічним законом Вебера-Фехнера***. Коротко закон формулюється так: однакові відносні зміни сили подразника викликають однакові абсолютні зміни відчуття, тобто відчуття пропорційне логарифму подражнюючої сили.

Слухове відчуття E визначається як

$$E = 10 \lg \frac{I}{I_N}, \qquad (2.2)$$

де I – інтенсивність звукового сигналу; I_N – поріг чутності на даній частоті.

Діапазон зміни слухового відчуття в області середніх частот від порога чутності до порога больового відчуття складає близько 130 дБ, що дорівнює величині стрибка відчуття 130/ 250 = 0,5 дБ. Стрибки відчуття залежать від частоти та інтенсивності сигналів. Мінімальна величина стрибка відчувається спостерігачем в області середніх частот і складає 0,4 дБ. В області низьких і високих частот за малих рівнів звукових сигналів стрибки відчуття доходять до 2… 3 дБ. Це свідчить про те, що слухове відчуття не зовсім збігається з законом Вебера-Фехнера. За частоти сигналу 1000 Гц з рівнем 80 дБ для подвоєння гучності потрібне збільшення сигналу на 10 дБ. Якщо сигнал дорівнює 20 дБ, то для подвоєння гучності достатньо збільшення сигналу на 5 дБ.

За одночасної дії на слух двох сигналів один з них може не прослуховуватись у присутності іншого, тобто спостерігається маскування одного сигналу іншим. Ефект маскування має дуже велике значення при розробці систем стиснення, тому що урахування маскування дозволяє суттєво знизити вимоги до апаратури і покращити її економічні показники.

Для тональних сигналів мірою маскування прийнято вважати величину підвищення порога сприйняття сигналу, який замаскований, над порогом його сприйняття у тиші. Кількісно маскування оцінюється в децибелах. При зміні порога чутності змінюється й рівень відчуття гучності сигналу.

Експериментально установлено, що низькочастотні сигнали сильніше маскують високочастотні. Коли частота завади більша, ніж частота корисного сигналу, то маскування менше.

Якщо сигнал маскується вузькосмуговим шумом, то при збільшенні смуги шуму маскування зростає лише до того часу, поки ширина смуги шуму не досягне деякого значення, яке називають критичною смугою слуху. У межах критичної смуги слуху вухо інтегрує інтенсивність заважаючих сигналів, а за межами смуги властивості інтегрування інтенсивностей вже не спостерігається. Ширина критичної смуги слуху залежить від середньої частоти смуги. В області низьких частот ширина критичної смуги складає десятки герц, а в області

високих частот – декілька сот герц. Слід зауважити, що результати визначення критичних смуг слуху, отримані різними авторами, суттєво відрізняються.

Коли сигнали, що сприймаються людиною, знаходяться в одній критичній смузі слуху, то вони взаємно маскують один одного. Якщо сигнали розмішені в різних критичних смугах слуху, то взаємне маскування менше, і сигнали сприймаються людиною майже незалежно. Ці якості слуху широко використовуються при розробці систем стиснення.

Широкосмугові звукові сигнали збуджують різні волокна, розміщені в мембрані вуха, а через слабку селективність слуху відбувається інтегрування спектральних складових сигналів у кожній з критичних смуг. Отже, вухо людини ніби перетворює суцільний спектр сигналів у дискретний, що має конкретне число складових, кількість яких дорівнює числу критичних смуг слуху.

Основні особливості маскування сигналу наглядно зображені на рис. 2.1, де показані криві порога чутності тонального сигналу при маскуванні його вузькосмуговим шумом з середньою частотою 1 кГц та полосою 160 Гц. На рис.2.1 зображено п′ять кривих порога чутності, що відповідають таким рівням вузькосмугових шумів: 100, 80, 60, 40, 20 дБ. Нижня крива відповідає порогу чутності тональних сигналів у тиші, тобто за відсутності вузькосмугового шуму (L – рівень сигналів, дБ; f – частота сигналів, кГц).

Найбільше підвищення порога чутності існує на середній частоті смуги шуму. Точки максимуму розміщені на 4 дБ нижче відповідних рівнів шуму, що пояснюється спроможністю слуху виділяти тональний сигнал на фоні шуму.

В області низьких частот криві спадають дуже круто, в області високих частот крутість менша. За великих рівнів шуму крутість спаду, особливо в області високих частот, суттєво зменшується. На частоті 1,7 кГц за рівня шуму 100 дБ спостерігається другий максимум. Однією з причин різкого збільшення маскувальної дії шуму за великих рівнів може бути нелінійність слуху [5].

Нелінійні властивості слуху проявляються за великих рівнів сигналів й установлені експериментально. Під час досліджень слухачам пропонували звуковий синусоїдальний сигнал. За малих і середніх рівнів цього сигналу не прослуховувались спотворення сигналу у вигляді суб′єктивного відчуття гармонік. Коли цей самий сигнал прослуховували за рівнів близько 100 дБ, то слухачі суб′єктивно відчували нелінійні спотворення у вигляді другої та третьої гармонік, хоча за допомогою приладів такі спотворення в сигналі не реєструвались.

При прослуховуванні двох тональних сигналів з великими рівнями суб′єктивно було чути гармоніки та комбінаційні частоти, тобто орган слуху людини за великих рівнів сигналів поводить себе як нелінійна система. Коли прослуховується мовний або музичний сигнали з великими рівнями, то відчуваються значні нелінійні спотворення [3].

Орган слуху людини інерційний. Слухове відчуття сигналу не зникає відразу після припинення дії джерела звуку, а поступово зменшується до нуля. Тому слабкі звуки, які йдуть після гучних, можуть бути повністю або частково замаскованими. При надходженні до слухача двох сигналів, які однакові за складом і рівнем, у випадку, коли один з них запізнюється за часом менше, ніж

на 50 мс, ці сигнали сприймаються суцільно. Якщо сигнал, що запізнюється, набагато менший сигналу, який надходить до слухача першим, то він не буде сприйматись роздільно навіть при запізненні більшому, ніж 50 мс.

Для звукових сигналів характерні швидке зростання і відносно повільне спадання рівнів сигналів. Мінімальний час зростання рівнів сигналів 3…5 мс, а час спадання – близько 0,5 с. Час спадання рівнів залежить не тільки від сигналу, але й від приміщення, в якому звучить сигнал, і може досягти декількох секунд. Середній час зростання рівнів мовних сигналів – 3…120 мс, а музичних сигналів – 20…140 мс.

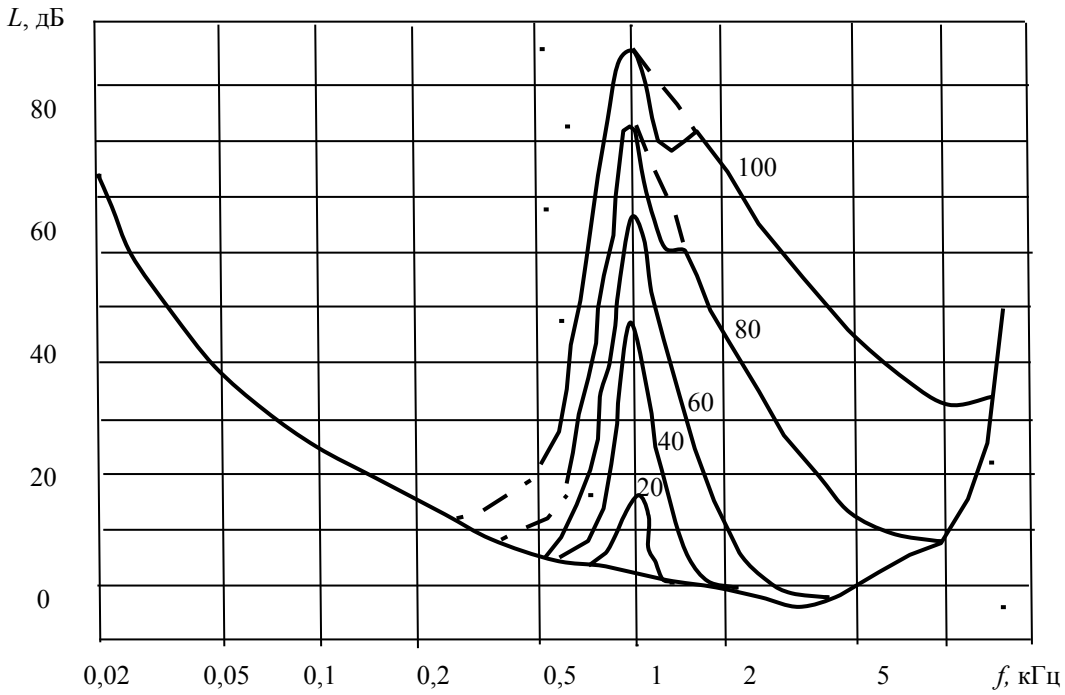


Рисунок 2.1 – Криві порога чутності тонального сигналу при маскуванні вузькосмуговим шумом з середньою частотою 1 кГц та полосою 160 Гц

У системах передачі звукових сигналів часто використовують паузи між сигналами. Паузи існують як при передачі мовних, так і музичних сигналів. Тривалість пауз – 50…3000 мс.

Контрольні питання

1. Що називають порогом чутності?
2. Дайте визначення поняття «інтенсивності звукового сигналу».
3. Чому погіршується слух людини при тривалому впливі гучних звуків?
4. В чому полягає дискретність слуху за амплітудою та частотою звукових сигналів?
5. Чим можна пояснити дискретність слухового сприйняття?
6. В чому полягає ефект маскування одного сигналу іншим?
7. Що називають критичною смугою слуху?
8. Коли проявляються нелінійні властивості слуху?
9. Як проявляється інерційність слуху людини?

3. МОВОУТВОРЕННЯ ТА РОЗБІРЛИВІСТЬ МОВИ

3.1. Теорія мовоутворення

Мова складається з фраз, слів, складів, звуків, але при всіх цих різноманітних мовних сигналах можна класифікувати близько 30…50 найменших звукових одиниць даної мови, які називають фонемами. При вимові кількість варіантів фонем в 3…5 разів перевищує кількість фонем, тому що варіанти фонем залежать від особливостей вимови. Розрізняють голосні та приголосні фонеми. Букви і фонеми можуть збігаютися, а можуть і відрізнятись. У різних мовах різна кількість фонем та різна кількість букв. В англійській мові існує 42 фонеми, а в російській – 41 фонема [6].

Багато приголосних мають тверду і м′яку форму, кожна з яких має свою фонему, тому що звучання твердих і м′яких приголосних дуже відрізняється. Наприклад, звучання твердої приголосної «з» та м′якої приголосної «зь» дуже різні.

Фонеми, які рідко зустрічаються у мові, несуть більше інформації, ніж ті фонеми, які зустрічаються часто. Голосні фонеми несуть меншу інформцію, ніж приголосні. Наприклад, якщо замість слова «посилка» вимовити звуки «о, и, а», а потім – «п, с, л, к», то в другому випадку легше здогадатися, яке це слово.

При вимові звуків мови в спектрі звукових сигналів існують ділянки, де спектральні складові максимальні. Ці ділянки визначають сприйняття і розпізнавання звуків мови. Джерела мовних коливань такі: гортань з голосовими зв′язками, вузькі щілини в мовному тракті.

Дзвінкі приголосні і голосні звуки мови утворюються голосовими зв′язками, що модулюють потік повітря, який проходить через них. Зв′язки не генерують коливань як струни, а в результаті своїх скорочень переривають потік повітря з періодом, обумовленим власною частотою коливань зв′язок. Середній період імпульсів повітря залежить від геометричних розмірів зв′язок і їх протяжності.

Мова людини має дискретний спектр. Основну частину коливань спектра голосу називають частотою основного тону (ОТ), яка обумовлена кількістю імпульсів потоку повітря крізь мовоутворюючий тракт людини. Період цих імпульсів називають періодом ОТ. В імпульсах, утворених голосовими зв′язками, міститься до 40 гармонік. Амплітуди гармонік ОТ зменшуються з підвищенням номерів гармонік. Частота основного тону низьких чоловічих голосів складає 40…70 Гц, а найвищих дитячих голосів – 400…500 Гц.

При вимові звуків «м, н, л, р» у спектрі існують тональні та шумові складові. Вимова звуків «ф, с, ш, х, т, к, ц, ч» супроводжується тільки шумовими складовими спектра, тобто голосові зв′язки майже не беруть участі в утворенні цих звуків, існують тільки незначні коливання зв′язок з частотою близько 30 Гц.

Шумові джерела звукових коливань знаходяться в місцях звуження мовного тракту або недалеко від нього. Шум створюється вихровими рухами потоку повітря, що проходить крізь щілини мовного тракту. При утворенні звуків

«п, т, к, б, г, д» з'являється імпульс шумового джерела, спектр якого має обвідну з крутістю, що дорівнює 6 дБ / октаву (при зміні частот спектра вдвічі, амплітуди спектральних складових також змінюються у два рази). Октавою називають інтервал частот, у якого верхня частота у два рази більша нижньої.

Мовоутворюючий тракт це трубка зі змінним перерізом, яку можна представити у вигляді системи, складеної із декількох резонаторів з зосередженими параметрами. Резонансні якості мовного тракту приводять до того, що деякі спектральні складові звуків будуть підсилюватись, а інші спектральні складові – послаблюватись. Резонансні частоти мовного тракту називають формантними частотами, або просто формантами. Частоти залежать від розмірів та форми мовного тракту [7].

Аналізуючи мовний тракт враховують вплив легенів, трахеї та гортані. Інколи мовний тракт представляють у вигляді послідовно з′єднаних секцій, що робить можливим точний розрахунок процесів мовоутворення. Але в цьому випадку важко визначити місцезнаходження формант у частотному діапазоні. Зображення мовного тракту у вигляді набору резонаторів менш точно характеризує мовний тракт, але таке зображення має більш наочний вид, тому частіше використовується при теоретичних дослідженнях і дозволяє отримувати необхідні аналітичні співвідношення. Процес формування мовних сигналів вивчається за допомогою акустичних та електричних моделей. Будь-яка форма мовного тракту може бути описана за допомогою формантних частот. Формантні частоти не залежать від частот гармонік ОТ.

Формантні частоти обумовлені резонансними об′ємами, існуючими у мовному тракті. Зміна розмірів резонаторів і щілин, які з′єднують ці об′єми, призводить до змін резонансних та формантних частот. Нумерують форманти у порядку зростання їх частот. Найбільш важливу роль при прийомі мовних сигналів відіграють перша та друга форманти. Третя і четверта форманти дозволяють розпізнати особу абонента. Співвідношення між формантними частотами та форма спектральних обвідних використовуються для розпізнавання звуків мови.

Аналіз обвідних спектрів голосних звуків показав, що в обвідних існує ряд додаткових піків та провалів. Також існують помилкові форманти, які не характеризують дані фонеми, але можуть бути розпізнавальною рисою особи абонента. Загальна інтенсивність мови визначається першою формантою. Для розпізнавання голосних звуків «а, о, у» достатньо першої форманти, а для розпізнавання звуків «е», «и» потрібні перша та друга форманти.

Приголосні звуки також можна визначити за формантними частотами та провалами у спектрі, але обвідні спектрів цих звуків мають немонотонну структуру, що ускладнює їх визначення. У мові ряд приголосних звуків ніби не мають власних спектрів, а являють собою зміну спектра сусіднього голосного звуку. Крім спектральної обвідної для розпізнавання приголосних звуків можна використовувати часові функції рівня мови.

Для розпізнавання звуків мови дуже важлива тривалість звучання. Тривалість голосних звуків залежить від того, у якому місці слова знаходяться ці звуки. У середньому тривалість наголошених голосних звуків складає 240 мс, а

ненаголошених голосних – 120 мс. Тривалість наголошеного звуку «а», в залежності від місця розміщення у слові, змінюється у межах від 180 до 260 мс, а ненаголошеного «а» – від 120 до 180 мс. Тривалість приголосних звуків в середньому дорівнює 95 мс.

Становить інтерес усереднений спектр мовних сигналів. Найбільший рівень мають спектральні складові в області середніх частот. При зниженні та при підвищенні частоти величина спектральних складових зменшується. Слух виявляє неточність передачі спектральних складових мови близько +3 дБ для піків та –9 дБ для провалів спектра. Якщо провали та піки спектра мови за шириною менші 1/8 октави, то вони не сприймаються слухом.

При підвищених емоціях рівень мови зростає на 3,5…6 дБ, а тривалість пауз зменшується на 15…25%. Підвищується також частота першої форманти на 5…12% і, відповідно, спектр усього сигналу зміщується угору за частотою. Якщо людина втомлена, то темп мови зменшується на 15…45% за рахунок збільшення тривалості пауз, знижується реакція на мову іншої людини, зменшується частота основного тону на 5…15%. При аварійній ситуації збільшується на 8…12 дБ рівень мови, а частота основного тону зростає в 1,4…2,1 рази.

У російській мові існує близько 3200 складів, кожний з яких має свою закономірність зміни формантних частот за часом та обвідну мови в цілому, що дозволяє виділити окремі склади під час аналізу і синтезу мови в вокодерах. Визначена відносна зустрічність слів з різною кількістю складів.

Середня довжина слова складає 2,9 складу, середня тривалість вимови складу – 250 мс, середня тривалість вимови слова – 800 мс. Аналіз вибірки з 80000 слів, сказаних під час телефоних розмов, показав, що в цій вибірці налічується всього 2240 слів. В загальному обсязі розмови всього 30 слів займають 50%, 155 слів – 80% та 737 слів – 96% [6].

При розробці апаратури треба враховувати сприйняття людиною спотворень та завад під час передачі мовних сигналів. Продукти нелінійних спотворень можна розглядати як завади, які корельовані з мовними сигналами. Оскільки продукти нелінійності, які знаходяться поза смугою пропускання тракту передачі, будуть подавлені, то основний вплив на мовні сигнали будуть чинити ті складові, які попадають в частотний діапазон мови.

Виявляється, що парні гармоніки мало впливають на розбірливість мови, а змінюють тільки тембр голосу. Більше значення мають непарні гармоніки, які викликають відчутні на слух спотворення у вигляді хриплого звучання. Комбінаційні складові продуктів нелінійності слабо відчутні, якщо вони збігаються за частотою зі спектральними складовими мовного сигналу, тому що відбувається маскування цих складових мовного тракту.

При передачі голосних звуків нелінійні спотворення мало відчутні, тому що, як правило, комбінаційні складові продуктів збігаються з початковими спектральними складовими і тільки трохи спотворюється обвідна спектрів.

При передачі приголосних звуків слухом сильніше сприймаються нелінійні спотворення.

Оцінку вірності сприйняття при прийомі фонем проводять за допомогою матриці порівняння, яка являє собою таблицю, до якої заносять результати роз-

пізнавання. Досьогодні не визначені ознаки фонем, за якими людина їх сприймає. Можливо, що в пам′яті людини закладені зразки слів з закінченнями та префіксами і за цими зразками відбувається сприйняття слів, а потім – окремих фонем. Автоматизм при прийомі слів пояснюється також тим, що відповідний образ слова створюється у мозку людини після надходження окремих фонем. Припускають, що в слуховому аналізаторі людини відбувається оцінка спектрограм і вибір згідно з ними того чи іншого звуку. Далі за окремими звуками складається образ слова.

З′ясування елементів сприйняття дозволило врахувати ці механізми при розробці фонемних вокодерів з високими якісними показниками. Поки що існують тільки гіпотези про дію механізмів сприйняття звуків людиною.

При розпізнаванні приголосних фонем використовують спектральні складові та часові характеристики мови. Однією з важливих часових характеристик є тривалість звучання звуку, тому що скорочення тривалості звучання приголосних звуків призводить до їх неправильного розпізнавання. Для правильного розпізнавання приголосних звуків знайдені оптимальні тривалості цих звуків. Так, наприклад, оптимальна тривалість звуку «х» складає близько 200 мс, а звуку «с» – близько 800 мс.

3.2. Теорія розбірливості мови

Визначальною характеристикою тракту передачі мовних сигналів є зрозумілість мови. З метою визначення зрозумілості мови розробили непрямий метод, що дозволяє кількісно визначити зрозумілість мови через її розбірливість.

Розбірливість оцінюється у відсотковій або відносній кількості правильно розпізнаних елементів мови. Такими елементами є склади, звуки, слова, фрази, цифри. Тому розрізняють складову, звукову, словесну, змістовну та цифрову розбірливості.

У ряді випадків при передачі мови ставлять вимогу розпізнавання голосу абонента, але визначальною характеристикою мовного тракту є розбірливість.

Для кількісного визначення розбірливості складені таблиці складів та слів, при розробці яких враховують, як часто зустрічаються ці елементи в даній мові.

У зв′язку з тим, що звуки, за винятком голосних, окремо не вимовляються, то при оцінці звукової розбірливості використовують таблиці звукосполучень та складові таблиці.

Оцінку розбірливості проводять спеціальні бригади, до яких входять молоді люди без дефектів мови та слуху. Ці бригади називають артикуляційними. Заздалегідь проводять необхідні тренування для отримання стійких результатів при одних і тих самих параметрах тракту. В процесі тренувань величина вимірюваної розбірливості буде зростати, тому що бригади поступово набувають необхідні навики та опановують методи вимірювань. Якщо не проводити тренування, то оцінки розбірливості трактів з однаковими параметрами будуть завищені або занижені.

Проведені дослідження по визначенню зрозумілості мови при звичайних телефонних розмовах. В дослідженнях брали участь люди різного віку, перевірялись тракти з різними параметрами. Розмови проводились в обидві сторони з використанням спеціальних розмовників.

Результати досліджень оцінювались за такою системою: «відмінно» – повна зрозумілість мови, розмови проводились без перепитувань; «добре» – були перепитування слів, які рідко зустрічаються, або прізвищ; «задовільно» – потрібні були часті перепитування, абоненти повідомляли, що важко розмовляти; «гранично допустимо» – потрібні неодноразові перепитування одного й того ж самого матеріалу, окремі слова передавали по буквах; «зрив зв'язку» – абоненти не розуміли один одного і відмовлялись від розмов. Для кожного з перевірених трактів з участю артикуляційних бригад були визначені показники розбірливості, які зведені в табл. 3.1.

Визначені статистичні залежності між складовою, словесною та змістовною розбірливістю мови. Результати, наведені в табл. 3.1, справедливі, коли передають та приймають на слух різноманітну інформацію. Якщо в обміні використовують обмежену кількість слів (наприклад, диспетчерський зв′язок), то відмінній зрозумілості мови відповідає менша складова розбірливість. При передачі цифрової інформації відмінна зрозумілість мови спостерігається навіть тоді, коли складова розбірливості складає лише 30%.

Таблиця 3.1 – Зрозумілість мови за різних градацій розбірливості

Зрозумілість	Розбірливість,%	
	Складова	Словесна
Гранично допустима	25-40	75-87
Задовільна	40-50	87-93
Добра	50-80	93-98
Відмінна	80 та вище	98 та вище

Зв′язок між умовами передачі сигналів та розбірливістю установлюють, враховуючи відчуття формант за наявністю шумів. Середня ймовірність появи формант на кожній ділянці частотного діапазону для конкретної мови має певне значення. Ймовірні характеристики розподілу формант на різних ділянках спектра досліджені тільки для деяких мов. Частотний діапазон мовних сигналів поділили на 20 смуг таким чином, щоб імовірність появи формант в кожній смузі була однакова і дорівнювала 0,05. Ці смуги називають смугами ***рівної розбірливості.*** В табл. 3.2 показані смуги рівної розбірливості для російської мови. Також наведені спектральні рівні мовних сигналів і шумів у цих смугах. У стовпчику 1 табл. 3.2 наведені мовні шуми у великому приміщенні з сумарним рівнем 65 дБ, у стовпчику 2 – мовні шуми у кімнаті з сумарним рівнем 60 дБ, у стовпчику 3 – шуми під час футбольного матчу з сумарним рівнем 60 дБ, у стовпчику 4 – промислові шуми з сумарним рівнем 77 дБ [7-8].

Положення формант в частотному діапазоні випадкове і залежить від вимовлених слів та особливостей людини. Основна енергія звуків мови зосереджена в формантах, тому рівні мови і рівні формант у смугах рівної розбірли-

вості майже збігаються. Шуми та завади зменшують рівень відчуття формант $E_ф$, який являє собою різницю між рівнями формант та маскувального шуму, тобто

$$E_ф = B_ф - B_ш = B_р - B_ш, \quad (3.1)$$

де $B_ф$ – рівень формант у смузі рівної розбірливості, який дорівнює за величиною середньому спектральному рівню мови $B_р$ у смузі рівної розбірливості.

Таблиця 3.2 – Смуги рівної розбірливості, спектральні рівні мовних сигналів і шумів

Номер смуги	Границі смуг рівної розбірливості, Гц	Ширина смуги, Гц	Серед-ня часто-та смуги, Гц	Спектральні рівні, дБ				
				Мови	Акустичних шумів			
					1	2	3	4
1	200-300	130	265	45,5	39,0	27,0	47,0	50,0
2	330-465	135	400	44,5	38,0	27,5	44,5	48,0
3	465-605	140	535	41,5	35,5	28,0	42,5	46,0
4	605-750	145	680	39,0	33,0	25,0	40,5	44,0
5	750-900	150	825	36,5	30,5	22,0	38,5	42,0
6	900-1060	160	980	34,0	28,0	19,0	36,5	41,0
7	1060-1230	170	1145	32,0	25,0	16,0	34,5	40,0
8	1230-1410	180	1320	30,0	23,0	13,0	32,5	39,0
9	1410-1600	190	1505	28,5	21,0	11,0	30,5	38,0
10	1600-1800	200	1700	27,0	20,0	9,0	28,5	37,0
11	1800-2020	220	1910	26,0	19,0	7,0	26,5	36,0
12	2020-2260	240	2140	25,0	18,0	6,0	24,5	35,0
13	2260-2530	270	2395	24,0	17,0	5,0	23,0	34,0
14	2530-2840	310	2585	22,0	16,0	4,0	21,5	33,0
15	2840-3200	360	3020	21,0	15,0	3,0	20,0	32,0
16	3200-3630	430	3415	20,0	14,0	2,0	17,5	31,5
17	3630-4150	520	3890	19,0	13,0	1,0	15,0	31,0
18	4150-4790	640	4370	18,0	11,0	0,0	12,0	30,5
19	4790-5640	850	5215	17,0	9,0	-2,0	9,0	30,0
20	5640-7000	1360	6320	15,5	7,0	-4,0	5,0	29,0

За результатами проведених дослідів установлено, що якщо в двох смугах рівної розбірливості забезпечується однакова ймовірність прийому формант, то внесок кожної з цих смуг у розбірливість мови однаковий. Для урахування зменшення ймовірності прийому формант вводять коефіцієнт розбірливості, який ще називають ***коефіцієнтом сприйняття***. Він являє собою ймовірність прийому формант.

Залежність коефіцієнта розбірливості ω_n від рівня відчуття формант показано в табл. 3.3.

Коефіцієнт розбірливості ω_n в кожній смузі різний. Сумарну ймовірність прийому формант A називають ***формантною розбірливістю***.

$$A = 0{,}05 \sum_{n=1}^{20} \omega_n . \qquad (3.2)$$

Таблиця 3.3 – Залежність коефіцієнта розбірливості від рівня відчуття формант

$E_ф$, дБ	ω_n, відн.од.	$E_ф$, дБ	ω_n, відн.од.	$E_ф$, дБ	ω_n, відн.од.
–12	0,01	–1	0,17	22	0,90
–11	0,015	0	0,20	23	0,915
–10	0,02	3	0,30	24	0,93
–9	0,03	6	0,40	25	0,945
–8	0,04	9	0,50	26	0,96
–7	0,05	12	0,60	27	0,97
–6	0,06	15	0,70	28	0,98
–5	0,075	18	0,80	29	0,985
–4	0,095	19	0,83	30	0,99
–3	0,11	20	0,86	33	0,995
–2	0,14	21	0,88	36	1,0

Експериментально знайдені залежності між формантною, складовою та словесною розбірливостями, вони показані в табл. 3.4.

Таблиця 3.4 – Залежність складової *S* та словесної *R* розбірливостей від розбірливості формант

A, відн.од.	*S*, %	*R*, %	*A*, відн.од.	*S*, %	*R*, %
0,05	5	30	0,55	84	98,5
0,10	15	63	0,60	87	98,8
0,15	26	76	0,65	90	99
0,20	36	85	0,70	92,5	99,2
0,25	46	90	0,75	95,2	99,4
0,30	54	93	0,80	96,5	99,6
0,35	62,5	94,5	0,85	98	99,7
0,40	69	96	0,90	99	99,8
0,45	75	97	0,95	99,5	99,9
0,50	80	98	1,00	100	100

При розрахунках розбірливості враховують спектральні характеристики шумів, параметри апаратури, яка застосовується в тракті передачі, та умови відтворення звукових сигналів.

Добре вивчені питання, пов′язані з маскуванням мовних сигналів рівномірними флуктуаційними шумами. Менше вивчений вплив імпульсних і вузькосмугових завад. При впливі широкосмугового рівномірного флуктуаційного шуму поріг чутності в критичній смузі слуху залежить від рівня спектральних складових шуму та ширини критичної смуги. Тональний сигнал буде прослуховуватись, якщо його інтенсивність дорівнює інтенсивності шуму в критичній смузі. Це можна пояснити високою селективністю слуху. Слух за своєю

селективністю відповідає гребінцю вузькосмугових фільтрів, смуги пропускання яких дорівнюють критичним смугам слуху.

Шуми, розміщені в інших критичних смугах, тим менше маскують сигнал, чим далі він знаходиться від даної критичної смуги частот. Сигнали, розміщені на більш високих частотах, ніж частота маскувального шуму, маскуються сильніше.

Контрольні питання

1. Які звукові одиниці називають фонемами?
2. Як можна представити мовоутворюючий тракт людини?
3. Дайте визначення формантних частот.
4. В області яких частот спектральні складові мови мають найбільший рівень?
5. Як оцінюється розбірливість мови?
6. Хто проводить оцінку розбірливості мови?
7. Дайте визначення смуги рівної розбірливості.
8. Дайте визначення коефіцієнта розбірливості.
9. Що називають формантною розбірливістю?

4. ОБРОБКА ЗВУКОВИХ СИГНАЛІВ

Основне завдання обробки сигналів – це зміна параметрів сигналу таким чином, щоб за обмежених технічних можливостей системи передачі зберегти при прийомі необхідну кількість відомостей про передане повідомлення. Повідомлення, які містяться в неперервному звуковому сигналі, визначаються тривалістю сигналу T, шириною частотного діапазону F та динамічним діапазоном D [6].

$$D = 10\lg\frac{I_{\max}}{I_{\min}}, \tag{4.1}$$

де $I_{\max}$, $I_{\min}$ – максимальна та мінімальна інтенсивності звукового сигналу.

Добуток

$$V = T\cdot F\cdot D \tag{4.2}$$

називають ***об'ємом сигналу***.

Пропускну здатність каналу зв′язку також можна виразити добутком трьох величин: тривалості роботи каналу, частотного та динамічного діапазонів.

Нижня границя динамічного діапазону каналу визначається рівнем завад, а верхня – величиною перевантаження каналу зв′язку. Пропускна здатність каналу зв′язку характеризує ту максимальну кількість відомостей, яку може пропустити канал. Якщо пропускна здатність каналу зв′язку перевищує об'єм сигналу, передані повідомлення не будуть спотворюватись.

При кодуванні можна змінити кожний з трьох параметрів сигналу за рахунок інших, але так, щоб об'єм сигналу залишився незмінним. Наприклад, можна збільшити ширину частотного діапазону, але зменшити динамічний діапазон. Якщо збільшити час передачі, то можна зменшити частотний та динамічний діапазони. При передачі мовних сигналів не всі три параметри рівноцінні. Наприклад, динамічний діапазон передає менше інформації про звуки мови, ніж часова функція сигналу.

Об'єм сигналу, який передають по каналу зв′язку, можна зменшити за рахунок компандування. Під час передачі параметр сигналу стискується компресором, а при прийомі відновлюється експандером. Пристрій, який складається з компресора та експандера, називають ***компандером***. В залежності від того, який параметр компандують, розпізнають компандування динамічного діапазону, частотного діапазону та часове.

Компандування буває безпосереднє та параметричне. За безпосереднього компандування не руйнується мікроструктура сигналу, а тільки змінюється її об'єм. При цьому сигнал може спотворюватись, але спотворення відіграють роль завад. З безпосередніх методів компандування в техніці зв′язку найбільш широко використовують компандування динамічного діапазону.

Параметричне компандування передбачає виділення із сигналу повільно змінних параметрів, для передачі яких потрібна значно менша пропускна здатність каналу зв′язку, ніж для самого сигналу. На приймальному кінці каналу

зв′язку з параметрів синтезується початковий сигнал, мікроструктура якого була зруйнована при передачі.

Одним із ефективних та простих методів обробки мовного сигналу є частотна корекція, яка підвищує розбірливість. Високочастотні складові спектра мовних сигналів мають значно менший рівень, ніж низькочастотні, хоча розбірливість, в основному, визначається високочастотними складовими. За наявності завад високочастотні складові маскуються сильніше, ніж низькочастотні. З метою підвищення розбірливості під час передачі збільшують рівень високочастотних складових спектра. При цьому всі спектральні складові мови майже однакові за величиною. Така мова не природна, дуже хрипла, але за наявності гладких флуктуаційних завад розбірливість мови дуже висока. Тому вже в самих мікрофонах, які використовуються для передачі мови, є частотна корекція з підйомом частотної характеристики в області високих частот. При використанні високоякісних мікрофонів з рівномірною частотною характеристикою підйом високих частот здійснюють у підсилювальному тракті.

4.1. Компандування та обмеження динамічного діапазону

Передача по каналу зв′язку сигналу з великим динамічним діапазоном часто неможлива, тому що динамічний діапазон каналу менший, ніж динамічний діапазон сигналу. Тому на вході каналу зв′язку встановлюють компресор, який стискує динамічний діапазон сигналу. Цей сигнал пропускають по каналу зв′язку, а потім за допомогою експандера розширюють динамічний діапазон до початкової величини.

При передачі мовних сигналів не в усіх випадках відновлюють початковий динамічний діапазон. Так, наприклад, у вокодерах стискують динамічний діапазон з метою скорочення об’єму параметрів, які передаються по каналу зв′язку. При прийомі динамічний діапазон сигналів не відновлюють.

Компресори та експандери являють собою автоматичні регулятори, коефіцієнти передачі яких змінюються під впливом сигналів, які проходять крізь них. Розрізнюють регулятори миттєвої та інерційної дії. В регуляторах миттєвої дії коефіцієнт передачі залежить від миттєвих величин звукових сигналів і регулювання майже безінерційне. Форма коливань на виході регулятора змінюється, тому з′являються значні нелінійні спотворення. В техніці звукового мовлення їх застосовують для захисту потужної апаратури від перевантаження по входу. Такі регулятори являють собою безінерційні обмежувачі амплітуд сигналів. Вони прості та надійні.

Інколи при передачі мови використовують граничне амплітудне обмеження, яке також реалізується за допомогою регуляторів миттєвої дії. Мовний сигнал при такому обмеженні являє собою послідовність прямокутних імпульсів постійної амплітуди, змінюються тільки інтервали між нульовими переходами. Коли виконується модуляція таким сигналом, то отримають практично телеграфний режим. При прийомі звуки мови будуть однакові та максимальні за рівнем. Мова

має добру розбірливість за наявності завад, тому що не маскуються звуки з малими рівнями. Якість такої мови низька, з′являються додаткові комбінаційні та гармонічні складові, які надають мові специфічного тембру.

Спектр амплітудно-обмеженої мови трохи змінюється, виникають додаткові частотні складові, які можна вважати завадами. Розбірливість мови буде більшою, якщо перед амплітудним обмеженням вирівняти частотний спектр.

Якість звучання і розбірливість амплітудно-обмеженої мови підвищується, якщо попередньо частотний спектр перемістити в область високих частот за допомогою амплітудної модуляції мовних сигналів. Амплітудне обмеження сигналів у більш високій області частот призводить до того, що гармоніки та комбінаційні частоти, як продукти нелінійних перетворень мовних сигналів, опиняються поза смугою мовного сигналу та легко відфільтровуються.

Якщо по додатковому каналу передавати часову обвідну мовного сигналу, яку виділили ще до амплітудного обмеження, а потім при прийомі відновити обвідну амплітудно–обмеженого сигналу, то розбірливість мови підвищується, а якість звучання майже не змінюється.

У схемах регуляторів інерційної дії коефіцієнтом передачі керує випрямлена та усереднена напруга сигналу. Нелінійні спотворення, які вносять такі регулятори, малі. Сучасні методи розрахунків автоматичних регуляторів дозволяють вибирати параметри схем таким чином, що нелінійні спотворення, які вносять регулятори, можуть заздалегідь задаватися та бути дуже малими [8].

Коефіцієнт передачі автоматичного регулятора змінюється за допомогою керуючого кола, в якому відбувається детектування та вирівнювання керуючої напруги. Тому зміни керуючої напруги запізнюються по відношенню до змін вхідних сигналів регуляторів. Керуюча напруга впливає на елемент, який змінює коефіцієнт передачі. Для таких цілей часто використовують транзистори, напівпровідникові діоди та підсилюючі каскади.

Згладжування напруги в керуючому колі регулятора проводять за допомогою зарядно-розрядного ланцюжка, який складається з конденсатора та резисторів. Конденсатор цього ланцюжка не може миттєво заряджатися і розряджатися, тому під час роботи інерційних регуляторів відбуваються перехідні процеси, які змінюють співвідношення між сусідніми звуками.

Перехідні процеси в компресорах призводять до того, що звуки з малим рівнем, які вимовляються після звуків з великим рівнем, будуть деякий час подавлені. Особливо це подавлення помітне при коротких звуках. Гучні звуки, що вимовляються за слабими, будуть мати короткочасні сплески. Якщо параметри зарядно–розрядного ланцюжка регуляторів підібрані правильно, то перехідні процеси слухом людини майже не сприймаються.

Експериментально доведено, що коли вибрати час заряду в межах 0,5… 2,0 мс, а час розряду – 0,15…2 с, то перехідні процеси майже не відчутні на слух. В компандерній системі для найменшої відчутності перехідних процесів часові параметри (час заряду і розряду) керуючих кіл компресора і експандера повинні бути попарно рівними. В цьому випадку відбувається майже повна компенсація спотворень, викликаних перехідними процесами. Присутність частотних та фазових спотворень в каналі зв′язку не дозволяє повністю усунути

спотворення, викликані перехідними процесами. Тому для захисту потужних підсилювачів від короткочасних перевищувань сигналів, які викликані перехідними процесами в автоматичних регуляторах, використовують обмежувачі амплітуд сигналів миттєвої дії.

Як самостійний прилад компресори (без експандера) застосовують з метою збільшення середнього рівня сигналу, це підвищує розбірливість мовних сигналів за наявності завад. Стиснення сигналу дозволяє без зміни потужності радіомовного передавача розширити зону обслуговування. Стиснення динамічного діапазону мовних сигналів у два рази еквівалентне збільшенню потужності передавача в 2,6 рази.

Стиснення мовних інформаційних програм звукового мовлення дозволяє при добрій якості мови мати достатню розбірливість за наявності завад. При цьому не спостерігаються характерні для граничного амплітудного обмеження спотворення (специфічний тембр через гармонічні та комбінаційні складові мовного сигналу).

При передачі мови використовують три варіанти компресорів: звукові, складові та за середнім рівнем. Розрізняються ці компресори тільки часом розряду керуючого кола. В звукових компресорах час розряду вибирають у межах 60…150 мс, в складових – 150…250 мс. У компресорах, призначених для регулювання середніх рівнів сигналів, час розряду керуючих ланцюгів складає 2…5 с.

Звуковий компресор зближує рівні голосних і приголосних звуків мовного сигналу. Динамічний діапазон сигналу на виході звукового компресора складає всього 10…15 дБ. Складовий компресор вирівнює склади мови, залишаючи майже незмінним співвідношення між рівнями голосних і приголосних звуків. Компресор за середнім рівнем мови вирівнює голоси абонентів, не змінюючи співвідношення між складами та звуками.

У вокодерах бажано застосовувати всі три типи компресорів, включаючи їх послідовно. Голоси абонентів спочатку вирівнюють за середнім рівнем, потім вирівнюють склади і на останньому етапі обробки стискують рівні звуків. Таке включення компресорів забезпечує найбільше стиснення динамічного діапазону мовних сигналів [6].

У компресорах, призначених для стиснення звуків мови, можна вибрати і менший час розряду керуючих кіл (приблизно 10…50 мс), тому що по каналу зв′язку передають не мовний сигнал, а тільки його параметри. При передачі параметрів мови аналізатор вокодера може і не реагувати на перехідні процеси в керуючих колах компресорів.

При компресії мовних сигналів малі складові збільшуються, що призводить до зростання розбірливості, але за дуже сильної компресії розбірливість різко падає через збільшення спотворень. Установлено, що стискати динамічний діапазон мовних сигналів доцільно тільки до 16…20 дБ.

Обробку звукових сигналів проводять також обмежувачами максимальних (підсилювачів-обмежувачів) та мінімальних (шумозаглушувачів) рівнів. Обмежувачі мінімальних рівнів запирають тракт передачі сигналів, якщо вхідний сигнал менший заздалегідь установленого рівня. Цей рівень вибирають таким чином, щоб не послабити найменші звукові сигнали. Шумозаглушувачі

дозволяють зменшити шуми в паузах передачі. Часові параметри керуючих кіл шумозаглушувачів вибирають такі: час заряду – 1…2 мс, час розряду – 100… 300 мс.

Обмежувач максимальних рівнів (підсилювач – обмежувач) за своєю дією схожий на автоматичне регулювання підсилення з затримкою. Якщо вхідний сигнал менший за певну задану величину, то схема працює як простий підсилювач. При перевищенні вхідними сигналами заданої величини коефіцієнт передачі підсилювача-обмежувача починає зменшуватись таким чином, що вихідні сигнали майже не зростають при збільшенні вхідних сигналів. У цьому випадку починається обмеження рівнів сигналів. При цьому не відбувається спотворення форм сигналів, а змінюється тільки коефіцієнт передачі, тобто не спостерігаються спотворення (обмеження) форм амплітуд великих сигналів. Обмежувач максимальних рівнів називають підсилювачем–обмежувачем тому, що він працює в двох режимах: підсилення й обмеження. Підсилювач-обмежувач є регулятором інерційної дії, зменшує рівні гучних сигналів та використовується для захисту апаратури від перевантажень. У вокодерах підсилювачі-обмежувачі використовують рідко, перевага надається компресорам.

4.1.1. Миттєве компандування

З метою зменшення швидкості цифрового потоку при передачі звукових сигналів використовують миттєве компандування. Як і за аналогового компандування, на передавальному кінці цифрової системи передачі підключають компресор, а на приймальному кінці – експандер.

Середня потужність шуму квантування за аналого-цифрового перетворення сигналу пропорційна ширині інтервалу квантування і не залежить від величини сигналу, тому при зменшенні рівня сигналу знижується відношення сигнал/шум квантування. Для того, щоб відношення сигнал/шум квантування не залежало від величини сигналу, потрібно змінювати крок квантування. Більшим сигналам повинні відповідати більші кроки квантування, а при зменшенні сигналу треба зменшувати кроки квантування. Для цього використовують миттєвий компресор, стала часу якого майже дорівнює нулю. Після компресора сигнал кодується із застосуванням рівномірних кроків квантування. Така послідовність операцій еквівалентна поділу сигналу на інтервали змінної ширини. В літературі таке квантування сигналу називають ***нерівномірним*** [9].

Постійне відношення сигнал/шум квантування забезпечує всім абонентам однакову якість звучання. Амплітудна характеристика компресора при цьому повинна бути логарифмічною, тобто

$$y = \lg x, \tag{4.3}$$

де

$$y = U_{\text{вих}} / U_{\text{вих. макс}}, \tag{4.4}$$

$$x = U_{\text{вх}} / U_{\text{вх. макс}}, \tag{4.5}$$

$U_{вх}$, $U_{вих}$ – миттєві значення вхідних і вихідних сигналів компресорів, а $U_{вх.\ макс}$ і $U_{вих.\ макс}$ – максимальні значення сигналів. Величини x і y – нормовані значення сигналів на вході та виході компресора, відповідно.

Значення x і y обмежені –1 та +1, причому для $x = 0$ відповідає $y = 0$, а для $x = \pm 1$ також $y = \pm 1$. Характеристика (4.3) не задовольняє цим умовам, тому що, при $y = 0$ значення $x = 1$.

Виконання логарифмічної залежності між вихідним та вхідним сигналами компресора забезпечить практично однакове слухове відчуття шумів квантування за різних величин звукових сигналів. Оскільки характеристика (4.3) не проходить через точки (0,0) та (1,1), використовують модифіковані логарифмічні характеристики, які відповідають умовам нормування, тобто $y = 0$ при $x = 0$, а $y = \pm 1$ відповідають значення $x = \pm 1$.

При стисненні звукових сигналів використовують квазілогарифмічну характеристику компресора типу μ

$$y = \frac{\lg(1 + \mu x)}{\lg(1 + \mu)} \quad \text{для } |x| \le 1 \tag{4.6}$$

та квазілогарифмічну характеристику типу A

$$y = \frac{Ax}{1 + \ln A} \quad \text{для } 0 \le x \le 1/A, \tag{4.7}$$

$$y = \frac{1 + \ln Ax}{1 + \ln A} \quad \text{для } 1/A \le x \le 1. \tag{4.8}$$

У формулах (4.6)…(4.8) параметрам μ та A надають конкретні числові значення. Чим більші величини μ і A, тим більше стиснення динамічного діапазону сигналів. Збільшення стиснення покращує співвідношення сигнал/шум квантування при малих сигналах та погіршує його при великих сигналах порівняно з рівномірним квантуванням.

Компресор з характеристикою типу A дає приблизно такий же виграш, як і компресор з характеристикою типу μ, якщо $\mu = A$. В області малих сигналів компресор типу A через лінійність характеристики (4.7) забезпечує трохи менший виграш порівняно з компресором типу μ.

У цифрових системах передачі використовують кусочно-лінійну апроксимацію характеристик, поділивши діапазон миттєвих значень сигналів на ряд сегментів. У кожному із сегментів амплітудна характеристика компресора апроксимується відрізком прямої лінії. На рис. 4.1 наведена позитивна гілка амплітудної характеристики компресора A – 87,6 / 13, яка складається з семи сегментів. Повна амплітудна характеристика має 13 сегментів, число A дорівнює 87,6.

У кожному із сегментів інтервали квантування однакові. В першому сегменті інтервал квантування Δ_0 найменший. В кожному іншому сегменті інтервали квантування також рівні між собою, але

$$\Delta_k = \Delta_0 2^{k-1}, \tag{4.9}$$

де Δ_k – інтервал квантування в сегменті з номером k.

У сьомому сегменті інтервал квантування максимальний

$$\Delta_{макс} = \Delta_0 2^6. \tag{4.10}$$

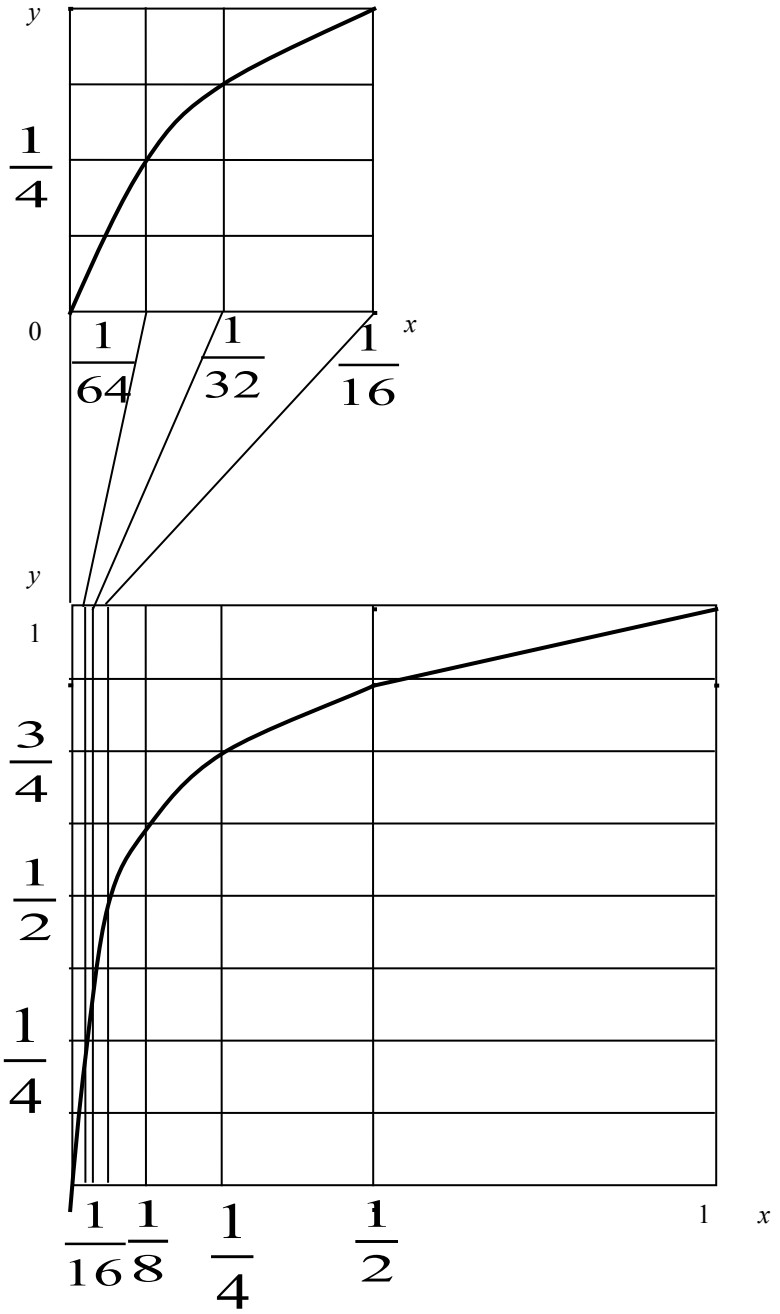


Рисунок 4.1 – Амплітудна характеристика компресора А-87,6/13

Малі сигнали, які знаходяться у першому сегменті, квантуються з розрізнавальною здатністю, що відповідає 14-розрядному рівномірному квантуванню. Кількість рівнів квантування в позитивній частині першого сегмента складає $2^8 = 256$. Кожний з наступних сегментів має 128 рівнів. При зростанні

номера сегмента інтервали квантування в старшому сегменті вдвічі більші, ніж у молодшому сегменті. Загальна кількість рівнів квантування складає 2^{10}=1024.

При використанні 14-розрядного рівномірного коду загальна кількість рівнів – 2^{14}=16384. Стиснення визначається, як зменшення кількості розрядів, потрібних для передачі одного відліку сигналу. В даному випадку величина стиснення буде 14 / 10 = 1,4.

У цифровій апаратурі передачі сигналів звукового мовлення також використовують компресор типу μ – 15 / 7. У цьому компресорі амплітудна характеристика має сім сегментів, μ = 15. Максимальна розрізнювальна здатність відповідає 11-розрядному рівномірному квантуванню. Великі відліки сигналу квантуються з розрізнювальною здатністю, яка відповідає 11-розрядному рівномірному квантуванню. Відношення максимального інтервалу квантування до мінімального – 2^3 = 8. Загальна кількість рівнів квантування 2^{12}, тобто для передачі треба мати 12–розрядний код, а стиснення дорівнює 14 / 12 = 1,17.

Амплітудна характеристика експандера має бути такою, щоб сумарна амплітудна характеристика була прямою лінією, що проходить через початок координат під кутом 45° до осі *x*. У цьому випадку компандер не буде вносити спотворень. Апаратуру з різними характеристиками компресії не можна з′єднувати безпосередньо, тому що наскрізна амплітудна характеристика компандера стане нелінійною і не буде відновлюватись вихідне співвідношення між амплітудами різних сигналів.

Якщо в каналі використовують неоднакові характеристики компресії, то відновити вихідні співвідношення між амплітудами сигналів можна двома засобами, а саме: 1) провести низькочастотний переприйом сигналів з необхідною корекцією сумарної амплітудної характеристики компандера; 2) здійснити транскодування, тобто цифрову переробку кодових слів компресора одного типу в кодові слова компресора іншого типу. В першому випадку погіршується якість передачі і відбувається підсумовування шумів квантування. При транскодуванні спотворення сигналів менші, ніж у випадку низькочастотного переприйому. Тому транскодування використовують частіше.

Відношення сигнал/шум квантування для компандера з характеристикою *A*-87,6/13 складає лише 50 дБ, а цього замало для високоякісної передачі сигналів звукового мовлення. Рівень шумів квантування повинен бути близько 60 дБ, щоб шуми були мало відчутні при слуховому сприйманні.

Компандер з характеристикою μ – 15/7 і рівнем шуму квантування близько 62 дБ майже задовольняє вимозі невідчутності шумів квантування. На жаль, ще й досі для передачі сигналів звукового мовлення використовують компандери, рівень шумів квантування яких не забезпечує невідчутності цих шумів вухом людини.

Якщо розглядати відчутність шумів квантування з тих самих позицій, що й відчутність продуктів нелінійних спотворень п′ятого порядку, то пороговий рівень їх відчутності буде близько 65 дБ. Отже, при рівні шумів квантування менше 65 дБ, можна бути впевненими, що вони не будуть відчутні при слуховому сприйнятті [9].

4.1.2. Майже миттєве компандування

Для стиснення динамічного діапазону сигналів часто використовують адаптивне квантування, за якого параметри квантування залежать від параметрів сигналу, виміряних за певний час. Один із методів адаптивного квантування, який називають майже миттєвим компандуванням, набув найбільшого поширення.

Кодові слова $X_{кв}(n)$, які отримані після рівномірного квантування та двійкового кодування звукового сигналу, надходять на лінію затримки ЛЗ місткістю М слів. Кожне кодове слово являє собою числове значення одного відліку звукового сигналу в часовій області. Для кожної групи з М слів керуюче кільце (КК) визначає масштаб квантування, який вибирають, виходячи з модуля максимального значення відліку в групі з М відліків (слів). 14-розрядні кодові слова $X_{кв}(n)$ за допомогою цифрового перетворювача (ЦП) перетворюються в 10-розрядні $X_{компр}(n)$, а далі надходять на пристрій об′єднання (ПО). Режим роботи цифрового перетворювача установлює керуюче кільце КК.

Для групи з М слів керуюче кільце визначає кодове слово $Z(n)$ коефіцієнта масштабу, яке надходить в пристрій об′єднання. На приймальному кінці каналу зв′язку пристрій розділення (ПР) виділяє 10-розрядні слова компресованого сигналу і кодові слова коефіцієнта масштабу. Цифровий експандер (ЦЕ) перетворює 10-розрядні слова в 14-розрядні $У_{кв}(n)$.

На рис. 4.2 показана схема майже миттєвого компресора, а на рис. 4.3 – його амплітудні характеристики.

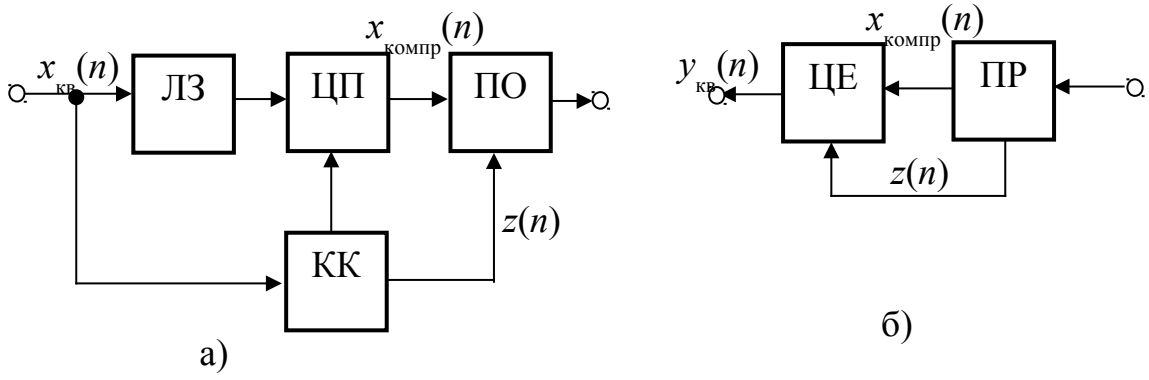


Рисунок 4.2 – Структурна схема майже миттєвого компандера:
а) компресор; б) експандер (ЛЗ – лінія затримки, ЦП – цифровий перетворювач, КК – керуюче кільце, ПО – пристрій об′єднання, ЦЕ – цифровий експандер, ПР – пристрій розділення)

Принцип перетворення кодових слів за допомогою миттєвого компандера можна пояснити табл. 4.1.

Таблиця 4.1 – Перетворення кодових слів майже миттєвим компандером

Модуль максимального відліку	Код на вході компресора	Код на виході компресора	Код на виході експандера
1>x>1 / 2	x1xxxxxxxxабсд	x1xxxxxxxx	x1xxxxxxxx0111
1 / 2>x>1 / 4	x01xxxxxxxабсд	x1xxxxxxxa	x01xxxxxxxa0111
1 / 4>x>1 / 8	x001xxxxxxабсд	x1xxxxxxаб	x001xxxxxxаб01
1 / 8>x>1 / 16	x0001xxxxxабсд	x1xxxxxабс	x0001xxxxxабс0
1 / 16>x>0	x0000xxxxxабсд	x0xxxxабсд	x0000xxxxxабсд

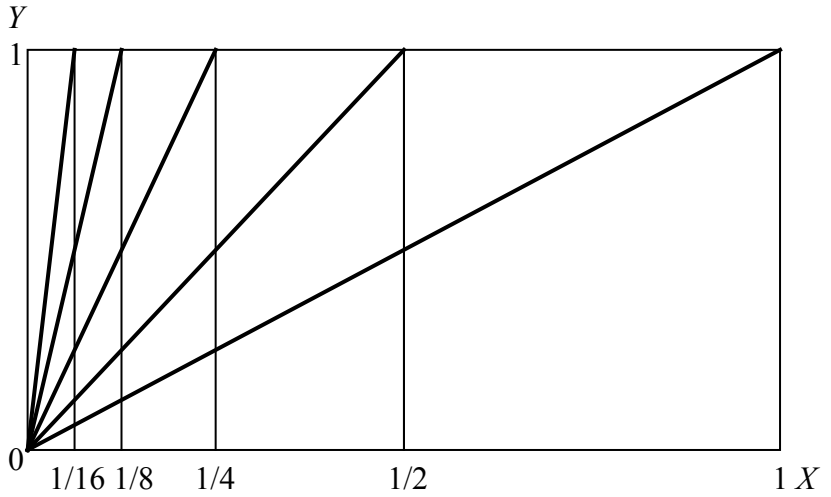


Рисунок 4.3 – Амплітудні характеристики майже миттєвого компресора

Амплітудні характеристики компресора (рис. 4.3) мають п′ять сегментів. X і Y нормовані значення вхідних та вихідних сигналів. Нормування сигналів відповідає формулам (4.4) і (4.5). Такі характеристики компресора дозволяють зменшити кодове слово на чотири розряди. Для відновлення М кодових слів на приймальному кінці каналу зв′язку треба знати величину коефіцієнта масштабу. Оскільки амплітудні характеристики компресора містять лише п′ять сегментів, то для передачі коефіцієнта масштабу достатньо трьох двійкових символів. Швидкість передачі цих символів мала, тому що кодове слово коефіцієнта масштабу передають тільки один раз у кожній групі з М відліків сигналу.

Коли максимальний нормований відлік сигналу в групі з М відліків не перевищує 1/16, то всі кодові слова цієї групи мають нулі в старших розрядах і їх можна не передавати на прийомний кінець каналу зв′язку, тому що ці нулі можна відновити за коефіцієнтом масштабу (див. нижній рядок табл. 4.1). Якщо максимальний нормований відлік в групі відповідає другому сегменту амплітудних характеристик компресора ($1/8 > x > 1/16$), то при передачі відкидається молодший розряд, це відповідає збільшенню інтервалу квантування вдвічі. На приймальному кінці числове значення відліку сигналу відновлюється за коефіцієнтом масштабу, але інформація про молодший розряд втрачається, тому в молодшому розряді записується 0 або 1.

Якщо робота компресора відповідає третьому, четвертому та п′ятому сегментам амплітудних характеристик, відкидаються, відповідно, два, три або чотири молодших розряди при передачі сигналу. При прийомі молодші розряди заповнюються 0 або 1.

Динамічні якості майже миттєвого компандера визначаються кількістю слів в групі, для яких передають один коефіцієнт масштабу. Цю кількість слів вибирають рівному цілому степеню двійки. За малої кількості слів у групі одержують невелике стиснення сигналу, а за великої кількості слів в групі можливі спотворення динамічного діапазону сигналів. Якщо в групі з M слів є одночасно відліки з малими і великими числовими значеннями, та оскільки коефіцієнт масштабу встановлюється, виходячи з найбільшого сигналу, то передача невеликих відліків здійснюється з дуже великими заокруглюваннями, що викликані відкиданням молодших розрядів. Тому тривалість групи слів за часом не повинна перевищувати часу зростання звукового сигналу. Не обов′язково враховувати час спадання звукового сигналу, бо він набагато більший часу зростання.

Мінімальний час зростання звукових сигналів складає близько 3 мс. За частоти дискретизації 32 кГц період дискретизації складає 0,031 мс. Тому кількість відліків у групі не повинна перевищувати 3/0,031 = 96.

Проведені суб′єктивно-статистичні експертизи підтвердили, що при $M \leq 32…64$ спотворення динамічного діапазону, які викликані роботою майже миттєвого компандера, не відчутні на слух. Якщо вибрати M = 32, то період передачі коефіцієнта масштабу складає 1 мс, це менше часу зростання сигналу, а швидкість передачі кодових слів коефіцієнта масштабу дорівнює 3 кбіт/с. Через те що на передачу кожного з M слів треба 320 кбіт/с, то загальна швидкість передачі сигналів звукового мовлення буде 323 кбіт/с. Стиснення сигналів, яке отримують при застосуванні майже миттєвого компандера, можна визначити як $(14 \times 32)/323 = 1,38$ [9].

Миттєвий компандер з характеристикою А – 87,6/13 забезпечує стиснення сигналів в 1,4 рази, тобто майже таке як і розглянутий в числовому прикладі майже миттєвий компандер. Як видно з рис. 4.1 та 4.3, динамічний діапазон сигналів з мінімальним інтервалом квантування у майже миттєвого компандера у чотири рази більший, ніж у компандера з характеристикою А – 87,6/13. Для цих двох компандерів передача сигналів у першому сегменті здійснюється з максимальною розрізнювальною здатністю, яка відповідає 14-розрядному рівномірному квантуванню. Отже, при передачі малих сигналів майже миттєвим компандером забезпечується краще співвідношення сигнал/шум квантування. Великі сигнали в майже миттєвому компандері квантуються з інтервалом у чотири рази меншим, ніж у миттєвому компандері з характеристикою А – 87,6/13. Таким чином, майже миттєвий компандер при передачі великих сигналів забезпечує більше співвідношення сигнал/шум квантування, ніж компандер з характеристикою А – 87,6/13. Майже миттєвий компандер рекомендований Міжнародним консультативним комітетом з телефонії і телеграфії (МККТТ) для передачі сигналів звукового мовлення [10].

Відношення сигнал/шум квантування при використанні майже миттєвого компандера залежить від кореляційних зв′язків між відліками сигналу, тому що використовується блок-пам′ять у вигляді лінії затримки. Точний розрахунок шумів квантування для майже миттєвого компандера складний, отримані тільки наближені співвідношення. Відношення сигнал/шум квантування для сигналів з великим рівнем у майже миттєвого компандера на 12 дБ більше, ніж у миттєвого компандера з тим самим стисненням сигналів. Це відношення може бути ще більшим, якщо поєднати в компресорі принципи майже миттєвого і миттєвого компандування, тобто при обробці групи кодових слів використати нерівномірне квантування та застосувати логарифмічні характеристики миттєвої компресії. Амплітудні характеристики нелінійного майже миттєвого компресора показані на рис. 4.4.

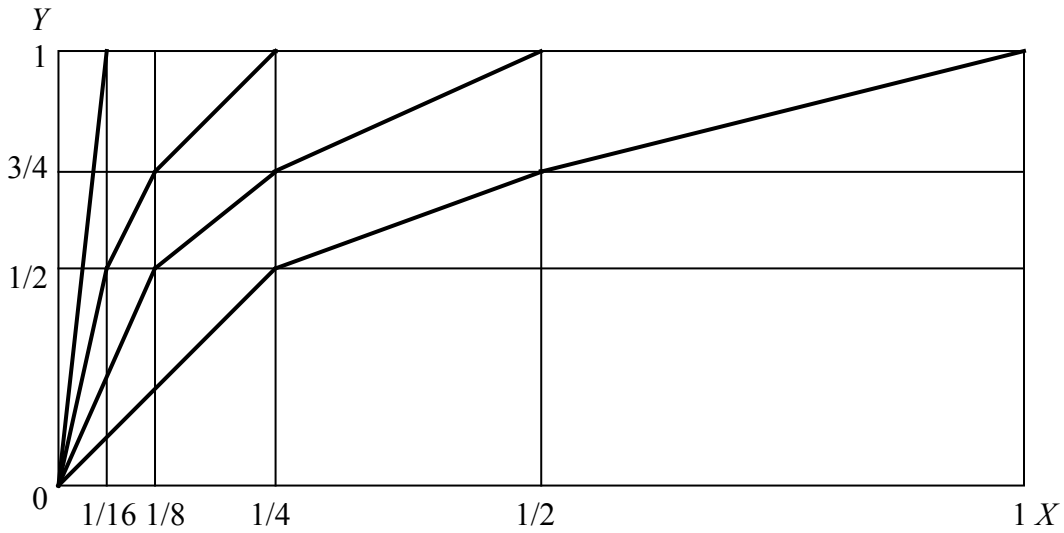


Рисунок 4.4 – Амплітудні характеристики нелінійного майже миттєвого компресора

4.2. Компандування та обмеження частотного діапазону

Були запропоновані два методи компандування частотного діапазону мовних сигналів. Суть методів полягала в діленні та множенні частотних спектрів. Методи показали помилковість у вихідних положеннях і не привели до позитивних результатів. Також пропонували ділити частотний діапазон на окремі смуги і передавати тільки частину кожної смуги. В цьому методі не враховувалось розподілення інформації про мову по частотному діапазону, передача була менш завадозахисною і метод був ефективним лише за відсутності завад та шумів.

Компресію частотного діапазону можна здійснити шляхом виділення миттєвої частоти і миттєвої амплітуди. В малому часовому відрізку сигналу міститься обмежена кількість частот з певними амплітудами. Потім можна організувати передачу цієї інформації по каналу зв′язку й отримати необхідне стиснення. Такі ж результати можна отримати, якщо використати перетворення

Фур′є у часовій та частотній областях сигналів. Докладно ці методи будуть описані у розд. 5.5.

Найпростіше реалізується скорочення частотного діапазону знизу і зверху. Якщо передача здійснюється за наявності флуктуаційних завад, то обмеження смуги мовних сигналів знизу частотою 300 Гц, а зверху частотою 3500 Гц підвищує розбірливість мови. Подальше обмеження смуги частот призводить до зниження розбірливості. Достатня розбірливість мови буде і при ширині смуги частот 1000 Гц, якщо в каналі зв′язку немає шумів та завад.

При проведенні суб′єктивно-статистичних експертиз установлено, що при скороченні смуги частот до 30…15000 Гц не більше ніж у 15% випадків помічається різниця у звучанні натурального та спотвореного сигналів. Менш помітно і сприятливіше для слуху одночасне скорочення в області низьких та високих частот. При обмеженні тільки низьких частот звук набуває відтінку, який дзвенить, при обмеженні тільки високих частот звук стає глухим.

Норми на обмеження смуги частот установлені експериментально на основі суб′єктивно-статистичних експертиз. Нормування параметрів якості здійснюється для таких смуг частот: 40…15000 Гц, 50…10000 Гц, 100…6400 Гц, 50…7000 Гц, 300…3400 Гц.

Для деяких трактів інколи допускають й інші смуги частот. Так, наприклад, в трактах проводового мовлення нормується смуга частот 100…6300 Гц.

4.3. Часове компандування

Відомо багато методів часового компандування. Ці методи використовують тільки для передачі мовних сигналів. При телефонній розмові двох абонентів кожний напрямок зайнятий лише на 50%, тому що коли один абонент говорить, інший його слухає. Якщо будуть говорити одночасно обидва абоненти, то вони не будуть розуміти один одного. Крім того, в самому мовному сигналі є паузи між окремими словами, фразами. ***Паузами*** називають нульові значення амплітуд звукового тиску, які існують при вимові звуків мови. Середній час тривалості пауз активної мови складає близько половини часу передачі. Тому активний час стану каналу зв′язку, коли по ньому передають мовні сигнали, складає близько 25% від усього часу телефонної розмови абонентів.

Найчастіше в мовному сигналі зустрічаються паузи тривалістю 50…150 мс, значно рідше появляються паузи тривалістю 2…3 с і більше. Паузи є і при передачі сигналів звукового мовлення. В табл. 4.2 показана кількість появ пауз та їх тривалість за одну годину передачі програми звукового мовлення. Результати отримані на підставі статистичних досліджень.

У табл. 4.2 також показана середня тривалість пауз у відсотках від загального часу передачі сигналів звукового мовлення [11].

З табл. 4.2 видно, що тривалість пауз складає всього 5% часу передачі, тому їх використання в сигналах звукового мовлення для передачі додаткової інформації навряд чи доцільне. В паузах передачі можна передавати коротко-

часні вимірювальні сигнали, за допомогою яких автоматично контролюють коефіцієнт передачі, амплітудно-частотні та нелінійні спотворення. Але, як правило, відмовляються від такої можливості, бо ці короткочастні сигнали будуть відчутні слухачам. З метою контролю раціональніше використовувати позначки точного часу або реальні сигнали звукового мовлення.

Таблиця 4.2 – Кількість появ пауз та їх тривалість за одну годину передачі програми звукового мовлення

Тривалість пауз, мс	Кількість появ пауз при передачі			Середня тривалість пауз,%
	мови	музики	програми у цілому	
50…150	414	74	215	0,6
150…300	148	27,6	79,3	0,5
300…500	130	28,2	73,4	0,8
500…1000	149	33,4	78,4	1,6
1000…2000	48	15,3	26,1	1,1
2000…3000	8	5,3	5,6	0,4
>3000	19	16,9	11,9	—
50…∞	913	200	490	5

Середня тривалість пауз мовних передач дорівнює 370 мс, а музичних – 800 мс. Ймовірність появ пауз та їх тривалість залежить від виду мови. Мова може бути активною (художнє читання без підготовленого тексту) та пасивною (читання підготовленого тексту). Середня тривалість пауз активної мови студентів близько 400 мс. У громадських діячів та письменників паузи мови триваліші і складають близько 710 мс.

При організації телефонних мереж зв′язку використовують статистичне ущільнення, яке дозволяє проводити під час пауз одного абонента передачу мовних сигналів інших абонентів. Такі системи передачі називають статистичними системами передачі.

У цих системах канал зв′язку надається абоненту тільки на час вимови його звуків. Якщо після паузи абонент знову починає говорити, то його підключають за допомогою швидкодіючих приладів комутації до іншого каналу зв′язку, вільного в цей час.

Системи передачі, які працюють за цим принципом, ефективні тільки за достатньо великих пучків каналів, коли час очікування зайнятого каналу не перевищує певного значення. Найчастіше такі принципи використовують в цифрових системах передачі, тому їх використання доцільне тоді, коли тракт передачі коштує дуже дорого. Їх використовують на трансатлантичних підводних кабельних магістралях та супутникових лініях зв′язку. Розроблена в США апаратура дозволяє при 100 каналах зв′язку обслуговувати 275 абонентів. При

пучках каналів порядку 1000 за рахунок статистичного ущільнення можна обслуговувати до 3000 абонентів.

На рис. 4.5 наведена, структурна схема системи передачі зі статистичним ущільненням, яка дозволяє об′єднати два потоки апаратури ІКМ-30 в один цифровий потік з такою ж самою швидкістю.

На рис. 4.5 показані передавач та приймач цифрової статистичної системи передачі (ЦССП) ІКМ-С2×30/30, передавачі та приймачі апаратури ІКМ-30, які дозволяють організувати односторонній зв′язок зліва направо. Для двостороннього зв′язку потрібні ще такий самий комплект апаратури та розв′язувальні пристрої, за допомогою яких можна з′єднати чотирипроводові закінчення двостороннього каналу магістральної системи з двопроводовою місцевою мережею.

Сигнали з передавачів ІКМ-30 надходять на визначники мови ВМ1 та ВМ2, які дозволяють визначити стан кожного вхідного каналу (активний чи пасивний). Абонент активного каналу в даний момент часу говорить, а абонент пасивного каналу – мовчить.

Визначники мови подають відповідні сигнали запиту в блок дисципліни обслуговування (БДО) активних каналів. Комутатор К і блок керування записом (БКЗ) забезпечують запис сигналів вхідних активних каналів в елементи пам′яті блока запам′ятовувального пристрою (БЗП). За кожним каналом лінійного тракту закріплені відповідні елементи пам′яті блока запам′ятовувального пристрою.

Однозначна відповідність вхідного активного каналу і наданим для передачі його сигналу вільним каналом лінійного тракту забезпечується за допомогою формувача сигналів керування (ФСК). Блок керування зчитуванням (БКЗч) формує цикл передачі, послідовно опитуючи елементи пам′яті блока запам′ятовувального пристрою.

Тактова частота і тривалість імпульсів, які забезпечують обробку та передачу інформації, задаються генераторним обладнанням (ГО). При прийомі дешифратор сигналів керування (ДСК) формує сигнали, по яких комутатор К відновлює просторове та часове положення тих сигналів, що надходять до абонентів. У приймачі ЦССП також є блок керування записом, блок керування зчитуванням і генераторне обладнання.

Блоки системи ІКМ-С2×30/30 реалізовані на мікросхемах. Недоліками системи є пропадання або спотворення звуків та складів, які викликані перевантаженням лінійного тракту та кінцевим часом спрацьовування визначників мови. Якщо кількість каналів лінійного тракту більша за кількість активних каналів, то спотворення викликаються тільки роботою визначників мови.

Якість передачі мовних сигналів за допомогою систем статистичного ущільнення відповідає нормам з розбірливості та відчутності спотворень. Спотворення, які проявляються у вигляді окремих клацань, мають таку саму інтенсивність, як і в каналах зв′язку, обладнаних звичайними системами з ІКМ.

В одному із методів часового компандування після вилучення пауз уривки мови, які залишаються, розтягуються на тривалість паузи до наступного

уривку мови. Це дозволяє зменшити частотний діапазон мовних сигналів, тобто часове компандування перетворюється в частотне.

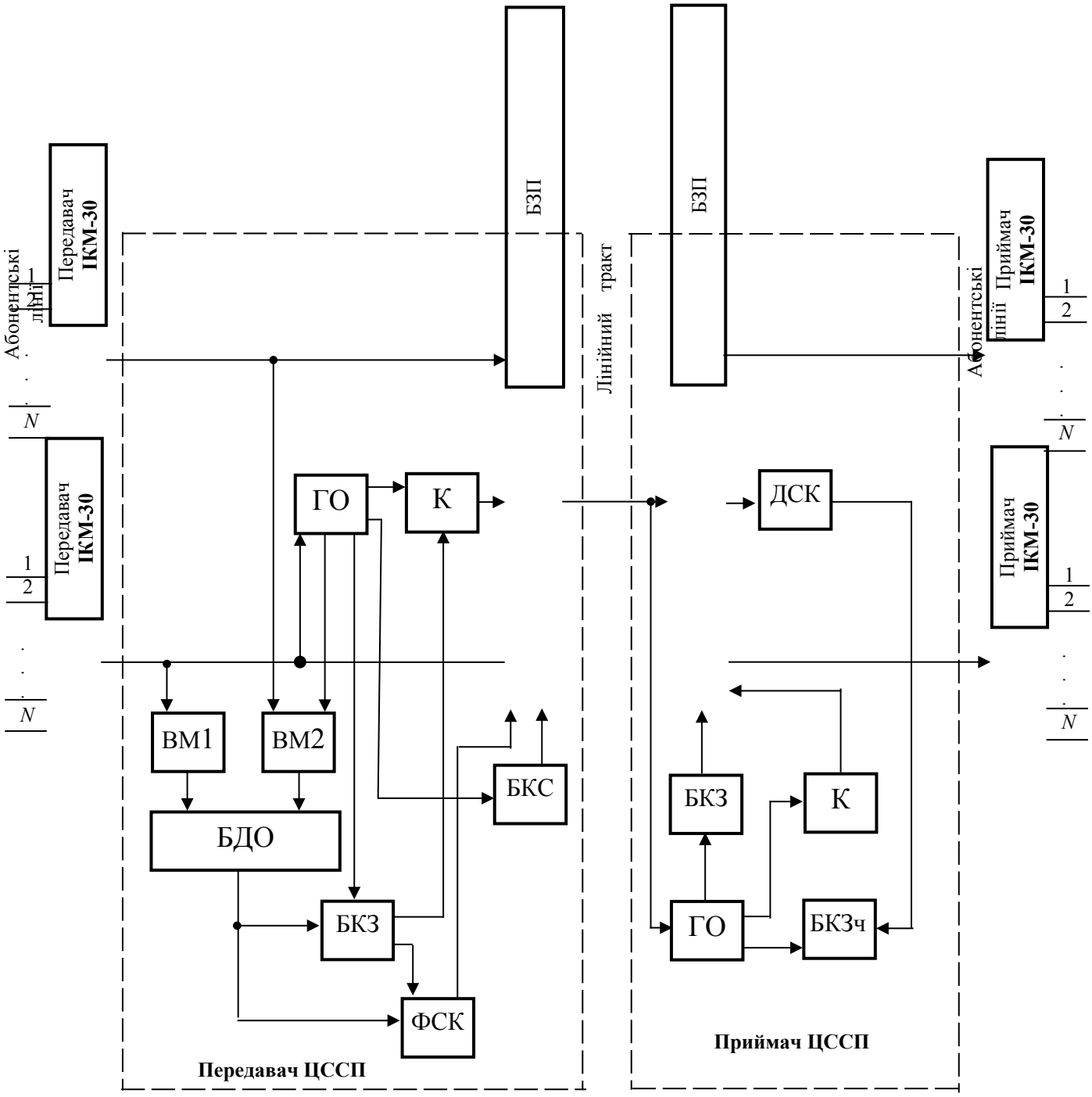


Рисунок 4.5 – Структурна схема цифрової передачі зі статистичним ущільненням:

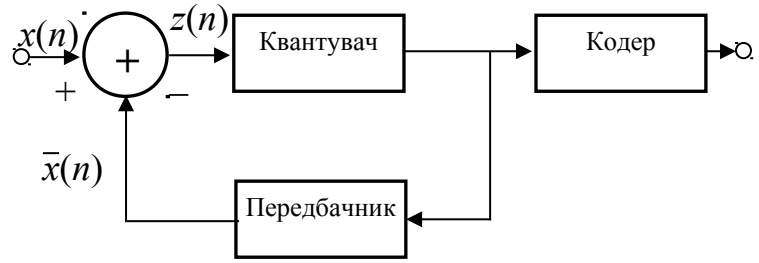
ВМ1, ВМ2 – визначники мови; ГО – генераторне обладнання; К – комутатор; БДО – блок дисципліни обслуговування; БКЗ – блок керування записом; ФСК – формувач сигналів керування; БКЗч – блок керування зчитуванням; БЗП – блок запам′ятовувального пристрою;

ДСК – дешифратор сигналів керування

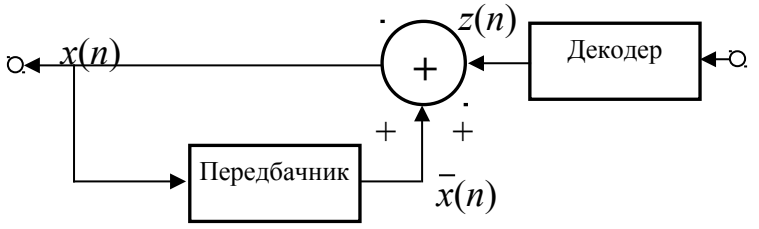
Голосні звуки і ряд приголосних мають відносно велику тривалість, при вимові цих звуків формантні частоти не змінюються, тому такі ділянки мови можуть вилучатись з передачі компресуючим пристроєм. На прийомі ці звуки можна відтворити за попередніми уривками цих самих звуків. Без зниження розбірливості та якості звучання можна виключити до 54% усієї тривалості мови. Отже, можлива двократна часова компресія мови, яка полягає у вилученні пауз і скороченні тривалості передачі окремих звуків.

4.4. Диференційні методи кодування

При диференційних методах кодування використовується кореляція між відліками сигналу, тому кодується тільки різниця між відліками. Для квантування різниці за однакової точності кодування треба менше порогів, ніж за квантування відліків. Структурна схема диференційного кодування показана на рис. 4.6.



а)



б)

Рисунок 4.6 Структурна схема диференційного кодування:
а) – передавальна частина; б) – приймальна частина

Послідовність відліків сигналу $x(n)$ надходить на один із входів віднімаючого пристрою передавальної частини, а на його інший вхід подають передбачені значення відліків $\bar{x}(n)$, що формуються передбачником. На вхід квантувача подають різницю

$$z(n) = x(n) - \bar{x}(n). \qquad (4.11)$$

Величину $z(n)$ називають помилкою передбачення, сигнали якої після квантування надходять на кодер передавальної частини. При прийомі сигнали $z(n)$ і $\bar{x}(n)$ надходять на суматор приймальної частини.

$$z(n) + \bar{x}(n) = x(n). \qquad (4.12)$$

Передбачники передавальної та приймальної частин однакові, тому в приймальній частині відбувається відновлення відліків сигналу $x(n)$ за формулою (4.12). Чим краще передбачник, тим менше помилка передбачення і менше порогів квантування необхідно для передачі сигналів $x(n)$ із заданою похибкою.

Сигнали помилки передбачення піддаються в цифрових системах передачі операціям квантування і кодування. Якщо для передачі сигналу похибки передбачення використовується імпульсно-кодова модуляція, то таке перетворення сигналу називають диференційною імпульсно-кодовою модуляцією (ДІКМ). Відомо багато варіантів технічної реалізації кодування з передбаченням. В одних системах сигнал помилки передбачення формується у аналоговій формі, а потім кодується, в інших системах спочатку кодують відліки сигналів, а потім формують сигнал похибки передбачення.

Уперше в світі питання передбачення випадкових послідовностей були розглянуті А. Н. Колмогоровим у роботі, надрукованій у 1941 році [12]. Через 8 років була надрукована робота Н. Вінера, теж присвячена цим питанням. До 1976 року було надруковано вже понад 700 робіт з адаптивних систем керування. Інтерес до цих питань не послаблюється й сьогодні з таких причин. По-перше, існуючі пристрої зображення сигналів у цифровій формі вимагають передачі значних об’ємів інформації із заданою швидкістю і точністю, що призводить до збільшення інформаційних потоків. По-друге, цифрові методи обробки сигналів дозволяють створювати прийнятні за техніко-економічними показниками системи з адаптивною обробкою і передбаченням сигналів, які дозволяють зменшити об’єм інформаційних потоків [13-19].

У звукових сигналах є надмірність, тому що існують кореляційні зв′язки між окремими відліками. Методи передбачення враховують кореляційні властивості звукових сигналів, а ефективність роботи передбачника оцінюється статистичними характеристиками похибки передбачення. Однією з таких характеристик є ентропія. Зі збільшенням кореляційних зв′язків між відліками сигналу ентропія похибки зменшується і прямує до нуля. В іншому граничному випадку, коли відліки сигналу статистично незалежні, ентропія похибки максимальна і дорівнює ентропії відліку.

Якість передбачення ще оцінюється такою статистичною характеристикою, як середньоквадратична помилка. Чим менша середньоквадратична помилка, тим кращий передбачник.

Аналітично та експериментально досліджено багато алгоритмів роботи передбачників. Порівняльний аналіз показує, що доцільніше застосовувати найпростіші передбачники. За ефективністю передбачення вони мало поступаються складним передбачникам і дуже прості при технічній реалізації [20-24].

При лінійному передбаченні числове значення передбаченого відліку подають у вигляді комбінації попередніх відліків

$$\bar{x}(n) = \sum_{k=1}^{M} \alpha_k x(n-k). \qquad (4.13)$$

Коефіцієнти α_k підбирають таким чином, щоб середньоквадратична помилка передбачення була мінімальною. Ефективність передбачення зростає із збільшенням кількості відліків M, за якими здійснюється передбачення. Якщо $M \to \infty$, то передбачення визначається спектром звукового сигналу. Доведено, що лінійне передбачення, здійснюване по всіх попередніх відліках, дозволяє зменшити динамічний діапазон сигналів не більше ніж на 12 дБ, що еквівалентно зменшенню двох розрядів при двійковому кодуванні. Динамічний діапазон сигналів симфонічного оркестру можна зменшити лише на 7,5 дБ, скоротивши всього один розряд при двійковому кодуванні. Якщо використовувати при передачі так звані критичні фрагменти програм, за яких найвідчутніші на слух спотворення, викликані роботою передбачників, то динамічний діапазон можна скоротити всього на 1,7 дБ.

Коли по каналу зв′язку передають критичні фрагменти програм звукового мовлення, то навіть зменшення одного розряду при двійковому кодуванні призводить до відчутних на слух спотворень сигналів. Тому недоцільно застосовувати лінійне передбачення за попередніми відліками сигналів в каналах звукового мовлення. Таке передбачення можна здійснювати тільки при передачі мовних сигналів у мережі телефонного зв'язку.

Якщо передбачення здійснювати за попереднім та наступним відліками сигналу, то помилка передбачення зменшується, а передбачений відлік розраховується як лінійна комбінація попередніх та наступних відліків

$$\bar{x}(n) = \sum_{k=-M}^{M} \alpha_k x(n-k). \qquad (4.14)$$

Як і раніше, передбачення зводиться до пошуку коефіцієнтів α_k, які забезпечують мінімум середньоквадратичної помилки передбачення.

Двостороннє передбачення, яке ще в літературі називають лінійним завбаченням «уперед» і «назад», забезпечує додаткове зменшення динамічного діапазону сигналів всього на 4 дБ без відчутних на слух спотворень.

У найпростішій системі з двостороннім передбаченням квантується різниця між даним відліком та половиною суми попереднього та наступного відліків. Коефіцієнти α_k в формулах (4.16) і (4.17) можуть бути оптимальними тільки для одного фрагмента звукового сигналу, для інших фрагментів ці коефіцієнти можуть суттєво відрізнятись.

У цифрових системах передачі використовують різні модифікації адаптивної диференційної імпульсно-кодової модуляції (АДІКМ). У цих системах крім передбачника значень сигналів міститься блок оцінки коефіцієнтів α_k. На приймальний кінець каналу зв′язку передають помилки передбачення і коефіцієнти α_k. При передачі з використанням АДІКМ виконують три перетворення: передбачення, оцінку коефіцієнтів і квантування. Адаптивне передбачення підвищує ефективність і якісні показники за рахунок використання змінних кореляційних властивостей сигналу.

При цифровій передачі сигналів звукового мовлення використовують майже миттєве компандування і ДІКМ. Відліки сигналу в цій системі розподіляються на парну та непарну послідовності. Парні відліки сигналу надхо-

дять відразу на майже миттєвий компресор, а непарні відліки спочатку підлягають диференційному квантуванню з двостороннім передбаченням і тільки потім надходять на майже миттєвий компресор. Така послідовність операцій забезпечує якісну передачу сигналів, тому що лінійне передбачення для цих сигналів неефективне. Розглянута система дозволяє організувати канал звукового мовлення при швидкості цифрового потоку 320 кбіт/с.

При зростанні частоти дискретизації збільшується кореляція між відліками сигналу, тому за великої частоти дискретизації число рівнів квантування передбачення можна зменшити до двох. Для передачі такої помилки вистачить одного розряду. Такий метод модуляції називають ***дельта-модуляцією*** (ДМ) [9].

Сигнал на виході дельта-модулятора показує тільки полярність сигналу помилки. Система з ДМ проста в реалізації. Сигнал помилки показує зменшився чи збільшився вхідний сигнал за час між двома сусідніми відліками. За допомогою рис. 4.7 пояснюється принцип формування сигналів при дельта-модуляції. На рис. 4.8 наведена структурна схема дельта-модулятора.

Неперервний сигнал $x(t)$ замінюється ступінчастою функцією $x'(t)$. Кожному кроку дискретизації відповідає приріст функції $x'(t)$ на величину, яка дорівнює кроку квантування. В передавальній частині дельта-модулятора формуються сигнали $z(t)$, які являють собою імпульси позитивної або негативної полярності. Якщо $(x(t) - x'(t)) > 0$, то формується імпульс з амплітудою, що дорівнює +1, а при $(x(t) - x'(t)) < 0$ – з амплітудою, що дорівнює –1.

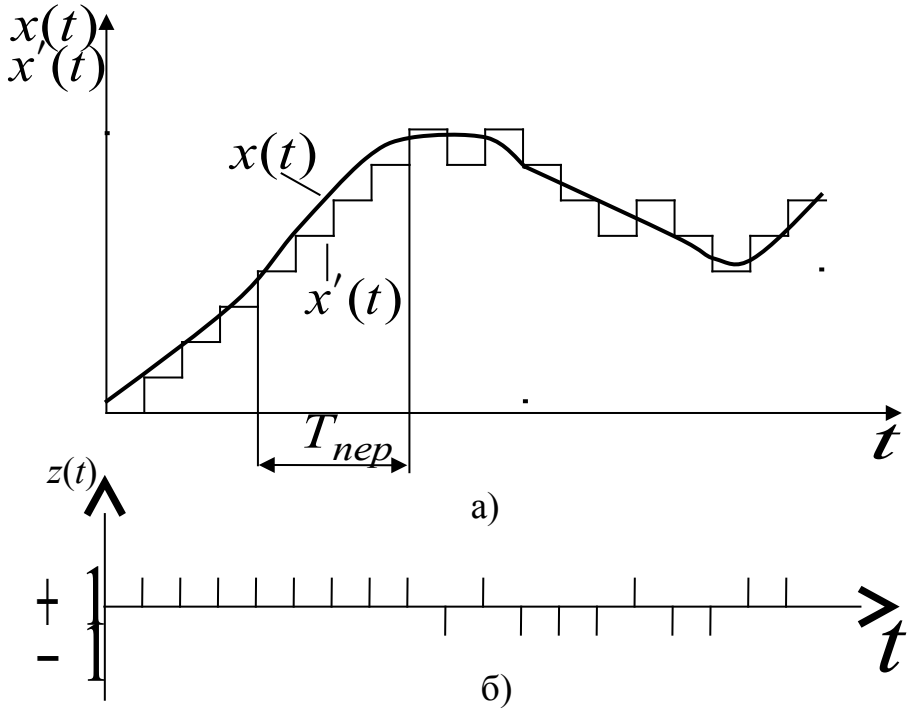


Рисунок 4.7 – До пояснення принципу дельта-модуляції

Імпульси з амплітудами +1 та –1 формуються на виході віднімального пристрою передавальної частини і надходять на однопороговий квантувач, роль якого може виконати компаратор К. Передбачником у схемі дельта-модулятора

є інтегратор І. Декодування сигналів у приймальній частині, тобто формування сигналу $x'(t)$, здійснюється інтегратором.

Частота дискретизації ДМ вибирається набагато більшою, ніж при звичайній ІКМ, тому що первинні сигнали відновлюються не за відліками, а за знаками фіксованих приростів амплітуд. На дільниці з великою крутістю сигналу спостерігається перевантаження кодера. На рис. 4.7 перевантаження кодера буде в інтервалі часу $T_{\text{пер}}$. Для того, щоб не перевантажувати кодер, треба так вибирати кроки квантування, щоб прирости сигналів $x(t)$ за час тактових інтервалів були не більші за кроки квантування.

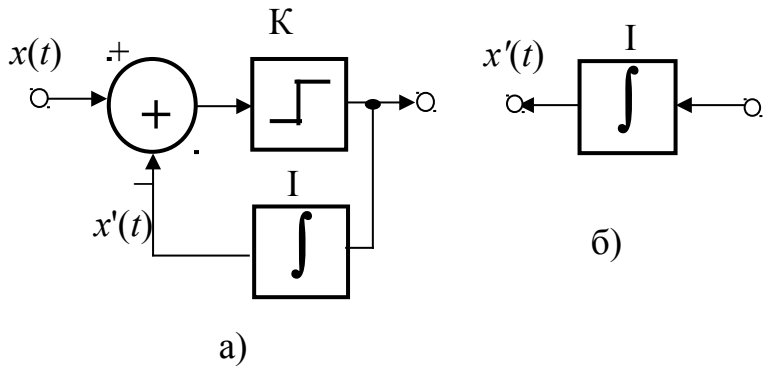


Рисунок 4.8 – Структурна схема дельта-модулятора:
а) – передавальна частина; б) – приймальна частина
(К – компаратор, І – інтегратор)

Для передачі телефонних сигналів тактова частота при ДМ повинна бути 150…200 кГц. Тактова частота при передачі телефонних сигналів за допомогою восьмирозрядної ІКМ дорівнює 64 кГц. Отже, при організації телефонних каналів з ДМ необхідна більша пропускна здатність, ніж при ІКМ. Другий із недоліків ДМ полягає в тому, що сигнали з малими і великими рівнями квантування квантуються з кроками одного розміру, що призводить до зайвої надмірності при передачі великих сигналів.

Зменшити частоту дискретизації при ДМ можна введенням змінного кроку квантування, який залежить від крутості вхідного сигналу $x(t)$. Цей алгоритм реалізований в адаптивній ДМ. Якщо зміна кроку квантування відбувається при переході від одного відліку до іншого, то така адаптація називається ***миттєвою***. Частіше застосовують адаптацію, за якої зміна кроку здійснюється один раз за певний інтервал часу.

У цифрових системах передачі застосовують адаптивну ДМ з неперервною зміною крутості, в якій використовують кероване компандування. Особливістю такої адаптивної ДМ є наявність пристрою, який дозволяє змінювати кроки квантування. При цьому можна суттєво знизити частоту дискретизації. Різні модифікації ДМ застосовують при передачі телефонних сигналів. Для передачі сигналів звукового мовлення ДМ майже не використовується через значні вимоги до якості передачі.

Контрольні питання

1. Що називають об’ємом сигналу?
2. Дайте визначення динамічного діапазону.
3. Що передбачає параметричне компадування?
4. Яка частотна корекція підвищує розбірливість мови?
5. Від чого залежить коефіцієнт передачі в регуляторах миттєвої дії?

6. Що керує коефіцієнтом передачі в регуляторах інерційної дії?
7. До чого приводять перехідні процеси в компресорах?
8. Як вибирають часові параметри керуючих кіл в компандерах?
9. В яких випадках використовують компресор як самостійний прилад?
10. Для чого використовують такі варіанти компресорів: звуковий, складовий та за середнім рівнем?
11. Для чого використовують обмежувач максимальних рівнів (підсилювач-обмежувач)?
12. З якою метою використовується миттєве компадування?
13. Для чого використовується транскодування?
14. Як вибирають масштаб квантування відліків сигналу при майже миттєвому компадуванні?
15. Що враховують при виборі кількості відліків у групі у разі майже миттєвого компадування?
16. Як сприймається на слух обмеження тільки низьких і тільки високих частот?
17. В чому полягає суть статистичного ущільнення?
18. Для чого використовують передбачники?
19. Що передають в канал зв’язку в системах диференційного кодування?
20. В чому полягає передбачення «уперед» і «назад»?
21. Поясніть принцип дельта-модуляції.

5. ВОКОДЕРИ

5.1. Принципи побудови

Компандування і диференційні методи кодування мовних сигналів дозволяють зменшити об’єм сигналу, але при цьому зберігаються всі внутрішні зв′язки, які мало впливають на розбірливість мови. В мовному сигналі за такої обробки залишається багато надмірної інформації, яка характеризує особу співрозмовника. В голосі людини міститься інформація про стан здоров’я, стать, вік, емоційний стан.

Якщо з мови вилучити надмірну інформацію, залишивши тільки ті параметри, які визначають розбірливість мови, то об’єм мовного сигналу скорочується в сотні разів.

Для передачі мови з доброю якістю при смузі сигналу до 7 кГц та динамічним діапазоном мови. який дорівнює 42 дБ, потрібна швидкість цифрового потоку до 100 кбіт/с, якщо використовується семирозрядний код та частота дискретизації 14 кГц. На ділянці від центральної нервової системи до мовного апарату людини швидкість передачі сигналів керування складає близько 100 біт/с. Така сама швидкість передачі необхідна при передачі мовних сигналів за допомогою телеграфного апарата. Причиною такої великої надмірності мови на виході артикуляційного апарату людини є спосіб створення мовного сигналу, за якого здійснюється спектрально-часова модуляція сигналу, який генерують джерела мовних коливань. Цей сигнал являє собою несучу частоту, яка характеризується в даний момент часу амплітудою, частотою і фазою. Параметри несучої частоти майже не містять необхідної інформації для розбірливості мови. Вузькосмуговий шум або гармонічний сигнал можна було б використати як несучу, тому що основна інформація про розбірливість мови міститься в спектрально-часових характеристиках. Якби виділити з мови тільки ті артикуляційні сигнали, які надходять з мовного центру мозку людини, передати їх по каналу зв′язку та промодулювати ними несучу, яка створюється тональними або шумовими джерелами, то для такої передачі треба було б 100 біт/с. На даному етапі розвитку науки і техніки поки що не вдається виділити з мови тільки корисну інформацію, яка містить відомості про розбірливість мови, тому організація мовного каналу зі швидкістю 100 біт/с лишається нездійсненною мрією людства [25].

Мовні сигнали передають за допомогою параметричних і мовоелементних вокодерів. В параметричних вокодерах при передачі виділяють параметри, які визначають зрозумілість мови й пізнавання голосу абонента, а при прийомі синтезують мовні сигнали, використовуючи параметри сигналу. Мовоелементні вокодери дозволяють при передачі розпізнати окремі елементи мови, а при прийомі синтезують ці елементи мови за певними правилами або вибирають їх з пам′яті приймального пристрою. Оскільки при передачі проводиться аналіз мовних сигналів, а при прийомі – їх синтез, то передавальні пристрої вокодерів називають ***аналізаторами***, а приймальні пристрої – ***синтезаторами***.

Аналізатор параметричного вокодера виділяє параметри, які характеризують спектрально-часову обвідну мовного сигналу і параметри джерел мовних коливань, які обумовлюють частоту основного тону і зміну її за часом. Визначають також момент появи та зникнення основного тону та шумового сигналу.

Синтезатор вокодера містить генератори основного тону і шуму, якими керують виділені при передачі параметри. Сигнали цих генераторів подаються на систему, що імітує роботу мовного тракту при вимові звуків. Сигнали генератора основного тону застосовуються при синтезі дзвінких звуків мови. При синтезі глухих звуків мови використовують генератор шуму.

Аналізатор містить пристрої обробки параметрів при поданні їх в канал зв′язку. В синтезаторі є пристрій розподілу параметрів і вторинної їх обробки.

За типом використовуваних параметрів розрізнюють смугові, гармонічні, формантні та фонемні вокодери. В смугових вокодерах при передачі визначають складові в вузьких смугах частот, у гармонічних – коефіцієнти Фур′є при зображенні сигналу у вигляді суми гармонік, у формантних – амплітуди і частоти формант, в фонемних – вимовлені звуки.

За конструкцією пристроїв виділення спектрально-часової обвідної мовного сигналу розрізнюють два види вокодерів: 1) з безпосереднім виділенням спектральної обвідної, яка змінюється за часом; 2) з використанням відліків сигналів у часовій області, за якими потім визначається спектрально-часова обвідна сигналу. Перший вид вокодерів називають ***спектральними***, а другий – ***спектрально-часовими***.

При передачі по каналу зв′язку параметрів в аналоговій формі за числом параметрів генеруються несучі частоти, кожна з яких модулюється своїм параметром. Несучі частоти розміщуються в смузі частот телефонного каналу. При передачі кожного з параметрів потрібна смуга близько 20…25 Гц. Число параметрів вибирається в межах 12…18 . Враховуючи необхідну розфільтровку сигналів параметрів, для їх передачі потрібна смуга частот близько 500 Гц, що дозволяє в смузі частот одного телефонного каналу розмістити до шести вокодерних каналів. На приймальному кінці каналу зв′язку перед поданням параметрів у синтезатор проводиться їхня розфільтровка і детектування. Аналогові вокодери при технічній реалізації стають дуже громіздкими, тому їх рідко використовують.

Цифрова обробка сигналів сприяла більш широкому впровадженню вокодерів, зменшенню їх габаритів і ціни. Часто використовують у вокодерах аналогову та цифрову обробку сигналів. Так, наприклад, виділяють параметри в аналоговій формі, а потім їх дискретизують, квантують і в цифровому вигляді передають по каналу зв′язку. При прийомі проводять перетворення параметрів з цифрової форми в аналогову, і тільки потім проводять синтез мовних сигналів, використовуючи параметри, отримані по каналу зв′язку [9].

У зв’язку з тим, що при передачі параметрів мовного сигналу в аналоговій формі необхідна смуга 20…25 Гц, то число відліків при дискретизації параметрів повинно бути в межах 40…50 за одну секунду. Збільшення числа відліків не призводить до помітного покращення якості і розбірливості мови. Кіль-

кість розрядів, виділяних для кодування амплітуд спектральних складових обвідної мовного сигналу, залежить від динамічного діапазону цих амплітуд.

Для передачі основного тону мови виділяють, як правило, до шести розрядів, для передачі спектральних складових обвідної мови – до чотирьох розрядів. Якщо передаються глухі звуки, то на передачу їх спектральних складових виділяють два розряди.

У вокодерах часто використовують такі швидкості передачі: 400, 600, 1200, 4800, 5300, 6300, 8000, 9600, 13000, 16000 та 32000 біт/с. В окремих випадках можуть використовутись й інші швидкості передачі. Так, наприклад, в мобільних телефонах використовується швидкість 14400 біт/с. Дослідженнями підтверджена велика надмірність вокодерних сигналів і наявність кореляційних зв′язків між параметрами, тому можна сподіватись на те, що швидкості передачі вокодерних сигналів будуть зменшуватись з розвитком науки і техніки.

При розробці й дослідженні вокодерних систем застосовують моделювання, яке дозволяє без виготовлення макетів апаратури отримати необхідні дані характеристик тих чи інших алгоритмів обробки сигналів. Моделюють як окремі блоки, так і вокодер в цілому. ЕОМ використовують як модель. Оцінку моделі проводять за допомогою артикуляційних таблиць або використовують стандартну фразу тривалістю до 15 с.

5.2. Аналіз та синтез мовних сигналів

Основний тон (ОТ) мови залежить від коливань голосових зв′язок при виголошенні дзвінких звуків мови. Параметри ОТ такі: частота та діапазон її можливих змін, амплітуда і мелодія. Мелодія являє собою усереднену за певний відрізок часу частоту ОТ. Статистичні характеристики параметрів ОТ залежать від характеру мовної інформації та голосів абонентів.

При дослідженнях статистичних характеристик параметрів ОТ використовують осцилограми коливань голосових зв′язок при вимові окремих фраз і звукосполучень. Виявляється, що частота імпульсів ОТ змінюється у великих межах. Швидкість зміни мелодії ОТ при переході голосний–приголосний і навпаки може бути до 6000 Гц/с. Тональні ділянки мови малі, основну частину мовного сигналу складають перехідні процеси. Під час коливання голосових зв′язок існують як невеликі флуктуації періодів ОТ, так і значні їх зміни, які можуть складати 30..50%. З імовірністю 0,85…0,95 відхилення тривалості суміжних періодів ОТ мови не перевищує 10%. Цікаві розподілення відношень пікових значень амплітуд ОТ в сусідніх періодах. З імовірністю 0,2 це відношення дорівнює 1, з імовірністю 0,8…0,9 – дорівнює 1,1, з імовірністю 0,7… 0,9 – дорівнює 1,3. Наведені дані свідчать про значні зміни амплітуд та інтервалів ОТ мовних сигналів.

Проведені дослідження частоти ОТ. На основі аналізу 335 телефонних розмов встановлено, що частота ОТ чоловічих голосів знаходиться у межах 70…180 Гц з середньою частотою 120 Гц, а жіночих – 180…330 Гц з середньою

частотою 240 Гц. Розподілення частот ОТ чоловічих і жіночих голосів підпорядковується нормальному закону. В літературі є й інші дані досліджень частот ОТ. Діапазон частот ОТ чоловічих голосів – 70–240 Гц при середній частоті 129 Гц, а діапазон частот жіночих голосів – 140…450 Гц при середній частоті 256 Гц. Різниця в результатах могла бути викликана різною кількістю досліджуваних голосів і характером мовних сигналів (дикторська мова, доповідь, телефонна розмова). Результати досліджень залежать і від віку людей, мова яких використовувалась при статистичній обробці. Частоти ОТ мови молодих людей вищі, ніж у літніх [6].

Розподілення тривалості шумових ділянок сигналу проводилось за осцилограмами чоловічих і жіночих голосів. Середня тривалість шумових ділянок мови (сигнал “шум”) складає 65 мс, а максимальні тривалості цих ділянок можуть бути 140…160 мс. Розподілення шумових і тональних ділянок підпорядковується логарифмічному закону.

Відчутність спотворень сигналів ОТ залежить від точності його виділення, а також від тих спотворень, які є при квантуванні й передачі сигналів по каналу зв′язку. Додаткові спотворення вносить і синтезатор. Для оцінки особливостей сприйняття сигналів ОТ і тон-шум були проведені дослідження з використанням 19-канального вокодера, в якому сигнали ОТ виділялись піковою схемою формування імпульсів ОТ. Наявності імпульсів ОТ відповідав сигнал “тон”, а їх відсутності – сигнал “шум”. Вокодер забезпечував складову розбірливість 94…96%.

Помилки в сигнал ОТ вводились або неперервно, або тільки на початку тональних ділянок. Середня тривалість ділянок мови з помилками складала 50 і 25 мс.

Виявилось, що при безперервних помилках відчутність спотворень складає 57%, коли вони складають 1% від усіх періодів ОТ. Якщо спотворені 3% періодів ОТ, то відчутність спотворень досягає 70%.

Трохи інакше сприймаються помилки, які вводяться на початку тональних ділянок. Якщо тривалість спотворень складала 25 мс, то при спотворенні 4% періодів ОТ це не відчувалась експертами. Тільки при спотворенні 12% періодів ОТ відчутність складала 63%, якщо тривалість спотворень була 25 мс. При тривалості спотворень 50 мс відчутність була рівною 70% [6].

Отже, відчутність спотворень ОТ спадає при зменшенні довжини початкової спотвореної ділянки. Тривалі спотворення сигналів ОТ добре сприймаються слухом.

Флуктуації позицій імпульсів ОТ сильніше відчуваються при передачі жіночих голосів. При передачі чоловічих голосів відчутність зміщень позицій імпульсів ОТ дорівнює 60%, якщо вони складають близько 90 мс, а при передачі жіночих голосів – близько 40 мс. Різниця у відчутності таких спотворень викликана тим, що середня частота ОТ жіночих голосів вища.

Якщо початкові ділянки дзвінких звуків оглушати шумом з тривалістю меншою ніж 10 мс, то слух не відчуває їх спотворень. Коли тривалість шуму на початкових ділянках дзвінких звуків досягає 20 мс, то відчутність таких спотворень досягає 90%.

Забезпечити велику точність при виділенні ОТ важко тому, що коливання голосових зв′язок мовного тракту квазіперіодичні. Одночасно зі зміною ОТ відбувається зміна форми голосового тракту. На ділянках мови з швидкими змінами мелодії ОТ форманти сильно змінюють структуру сигналу голосових зв′язок, тому важко виділити період ОТ. Виділення ОТ за піками і за переходами сигналу через нуль часто дає різні значення частоти ОТ.

Це пояснюється тим, що на положення піків впливають форманти тональних ділянок сигналу, а позиції переходів через нуль визначаються і формантами, і шумами. Якщо мова має малий рівень, то важко розрізнити шумові і тональні ділянки.

Умовно виділячі основного тону (ВОТ) можна розподілити на три групи: ВОТ, які використовують часові властивості мовних сигналів, ВОТ на основі частотних властивостей сигналів та комбіновані. Часові ВОТ використовують інтервали часу між піками і переходами через нуль кореляційної функції сигналу. ВОТ на основі частотних властивостей використовують періодичність сигналів за частотою. В комбінованих ВОТ вирівнюють амплітуди частотного спектра сигналу і за характером кореляційної функції визначають ОТ.

В схемах ВОТ використовують лінійне передбачення. Помилка передбачення має різкі викиди в кожному періоді ОТ і за цими викидами виділяють ОТ.

Піки сигналу можна виділити за допомогою порогових пристроїв, в якості яких часто використовують тригери Шмідта. Щоб виділити піки негативної і позитиної частин сигналу мовного тракту, використовують два тригера Шмідта. Якщо в мовному сигналі подавлені частоти нижче 400 Гц, то виділити ОТ важко, тому що рівень першої гармоніки ОТ невеликий. При цьому виділення ОТ за піками дає до 15…20% помилок. Скорочення низькочастотних складових мовного сигналу зменшує число помилок при виділенні ОТ в 3…4 рази.

ОТ можна виділяти за допомогою фільтрів, настроєних на першу гармоніку ОТ. Смуга пропускання фільтрів може бути регульованою. В паузах сигналу і при глухих звуках мови настроювання фільтра зберігається.

Дефектом таких схем є наявність помилок при швидких змінах ОТ. За рахунок згладження флуктуацій ОТ можливі великі фазові спотворення. Помилки при виділенні ОТ можуть виникати і під впливом шумів, якщо частота ОТ визначається за числом переходів сигналу через нуль.

Застосовують ВОТ, які складаються з декількох елементарних ВОТ. Обробка сигналу проводиться одночасно в декількох каналах, де аналізуються максимальні значення сигналів в певній смузі, різниця між максимальними та мінімальними значеннями, різниця між сусідніми максимальними значеннями, різниця між сусідніми мінімальними значеннями. Для аналізу можуть бути використані й інші параметри сигналу. Інформація з усіх паралельних каналів надходить на блок статистичної обробки та прийняття рішення. Мовний сигнал на виході синтезатора, до складу якого входить такий ВОТ, звучить природно і зберігає основні властивості голосу абонента.

Більш прості схеми ВОТ мають на вході декілька смугових фільтрів, що перекривають діапазон можливих значень частот ОТ мовних сигналів. До кожного смугового фільтра підключають піковий виділяч ОТ з пороговим

пристроєм, рівень спрацювання якого вибирають більшим рівня шумових сигналів. Коли частота першої гармоніки ОТ потрапляє в смугу пропускання цього фільтра, то на виході порогового пристрою в даному каналі з′являється сигнал ОТ, частоту якого можна визначити за числом переходів сигналу через нуль за вказаний інтервал часу.

Кореляційні ВОТ використовують періодичність кореляційної функції. В них є лінія затримки з відводами. Час затримки сигналу до першого відводу повинен відповідати мінімальній тривалості періоду ОТ, а час затримки всієї лінії – максимальній тривалості періоду ОТ.

Перший максимум кореляційної функції $B(\tau)$ сигналу

$$B(\tau) = \frac{1}{T} \int_{t-T}^{t} f(t) f(t - \tau) d\tau \qquad (5.1)$$

відповідає $\tau = 0$, а другий максимум $-\tau = \tau_0$, тобто величина τ_0 дорівнює тривалості періоду основного тону мови.

У формулі (5.1) величина T дорівнює максимальному часу затримки, а $f(t - \tau)$ і $f(t)$ затримане на відрізок часу, який дорівнює τ, поточне значення мовного сигналу. Мінімальну величину затримки вибирають близько 2 мс, а максимальну величину –14 мс. Для отримання необхідної точності визначення частоти ОТ число відводів лінії затримки повинно бути близько 140.

Запропоновані схеми ВОТ на основі кепстрального аналізу. В цих ВОТ за допомогою дискретного перетворення Фур′є визначається миттєвий спектр сигналу, потім визначають логарифм комплексного спектра потужності сигналу і обернене перетворення Фур′є цього логарифма. Обернене перетворення Фур′є логарифма комплексного спектра потужності називають ***кепстром***. Кепстр являє собою функцію, аргументом якої є час, а не частота. Кепстр має гострий максимум, якщо аргумент дорівнює за величиною періоду ОТ мови. Це і дозволяє використовувати кепстр для визначення періоду ОТ.

Схеми виділячів тон-шум (ВТШ) призначені для подачі сигналів «тон» і «шум» в канал зв′язку. Ці сигнали застосовуються синтезаторами вокодерів при відновленні звуків мови. ВТШ використовують в якості ознак звуків мови, характера розподілення спектральних складових за частотним діапазоном і періодичністю тональних ділянок мовного сигналу. Сигнал «тон» відповідає дзвінким звукам мови, а сигнал «шум» – глухим.

Основні ознаки відмінності дзвінких і глухих звуків мови такі: 1) енергія дзвінких звуків зосереджена переважно в області низьких частот, а енергія глухих звуків – в області високих частот; 2) енергія дзвінких звуків розподілена в області формантних частот, а енергія глухих звуків розподілена за спектром більш рівномірно; 3) дзвінкі звуки являють собою періодичні коливання, а глухі звуки – шумові коливання; 4) енергія глухих звуків рівномірно розподілена за часом, а енергія дзвінких звуків пульсує з частотою ОТ мови.

Мовний сигнал на тональних ділянках має у визначеній смузі більший рівень, ніж на шумових ділянках, тому обранням певного порога можна відділити сигнали «тон» і «шум». При розмові різних абонентів треба змінювати поріг спрацювання, оскільки рівень мови змінюється, тому такі ВТШ рідко використовуються. Якщо проводити аналіз сигналу на різних ділянках спектра,

то ефективність роботи ВТШ підвищується. Так, наприклад, схема ВТШ може бути об′єднана із схемою аналізатора спектра в смуговому вокодері.

Оскільки ВОТ працює і при передачі шумових ділянок мови, то можна аналізувати сигнали на виході ВОТ з метою визначення ділянок з періодичними послідовностями (сигнал «тон»). Якщо період імпульсів на виході ВОТ буде змінюватись хаотично, то це відповідає сигналу «шум». Періодичність мовних сигналів може аналізуватись і кореляційним ВОТ.

Схеми ВТШ можуть використовувати одночасно декілька ознак, що забезпечує краще виділення сигналів «тон» і «шум». Доцільно об′єднувати схеми ВОТ і ВТШ, це дозволяє спростити схеми вокодерів.

Моделлю мовоутворюючого тракту людини є чотириполюсник, який збуджується квазіперіодичними імпульсами при дзвінких звуках та – шумовими сигналами при глухих звуках. Цій моделі відповідає схема формування сигналів збудження, яка складається з керованих генератора імпульсів та генератора шуму. За наявності сигналу «тон» звук синтезується за допомогою генератора імпульсів. Якщо в даний момент часу надходить сигнал «шум», то для синтезу звуку використовується генератор шуму.

Частотою генератора імпульсів керують сигнали ОТ мови. Технічно найпростіше реалізується керований генератор прямокутних імпульсів, хоча форма імпульсів голосових зв′язок близька до трикутної. Тому сигнал на виході синтезатора вокодера не зберігає природності звучання.

Якщо при синтезі використовувати сигнали натуральної мови, то помітно покращується якість звучання чоловічих та жіночих голосів. Пристрої, в яких замість сигналів ОТ і тон-шум передають в кожній смузі натуральну мову, називають ***напіввокодерами***. Навіть при передачі натуральної мови в смузі 900 Гц, отримують відмінну якість звучання. У напіввокодерах з мови виділяється ще інформація про сигнали ОТ і тон-шум, які використовуються при синтезі. Звичайно, що напіввокодери забезпечують менше стиснення мовного сигналу порівняно з вокодерами.

5.3. Спектрально-смугові методи

На рис. 5.1 показана структурна схема смугового вокодера. Мовний сигнал надходить на вузькосмугові фільтри СФ1…СФ*N* і на ГОТ, ВОТ, ЕК та ГШ вузькі смуги.

У кожній частотній смузі відбувається детектування сигналу за допомогою амплітудних детекторів АД1…АД*N*. Потім сигнали надходять на фільтри нижніх частот ФНЧ1…ФНЧ*N*. Детектори і фільтри нижніх частот дозволяють виділити обвідні сигналів у кожній частотній смузі. Сигнали на виходах фільтрів пропорційні середнім значенням сигналу в кожній з частотних смуг. Інформація з фільтрів нижніх частот надходить на кодер аналізатора. При передачі виділяється частота основного тону мови за допомогою ВОТ. Ознаки звуків мови оцінюються ВТШ. Сигнали ВОТ і ВТШ також подаються на вхід кодера, з якого інформація надходить до каналу зв′язку в цифровій або аналоговій формі.

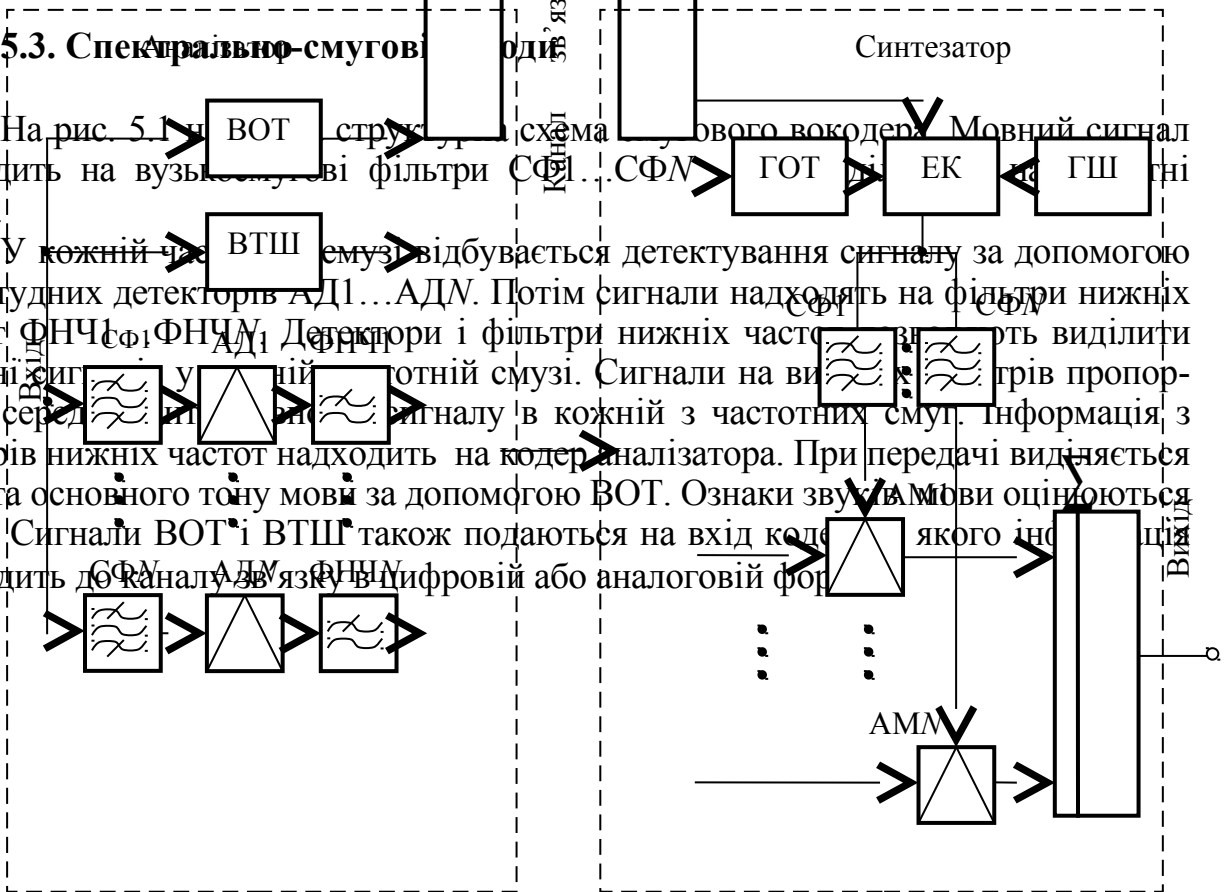


Рисунок 5.1 – Смуговий вокодер:
ВОТ – виділяч основного тону; ВТШ – виділяч тон-шум; СФ1…СФ*N* – смугові фільтри;
АД1…АД*N* – амплітудні детектори; ФНЧ1…ФНЧ*N* – фільтри нижніх частот;
ГОТ – генератор основного тону; ЕК – електронний ключ; ГШ – генератор шуму;
∑ – суматор; АМ1…АМ*N* – амплітудні модулятори

Сигнали, передані по каналу зв′язку, виділяються декодером синтезатора. Сигнали ОТ керують частотою генератора основного тону (ГОТ), який створює широкосмугові імпульсні сигнали. Електронний ключ (ЕК) подає на входи смугових фільтрів (СФ1…СФ*N*) синтезатора сигнали збудження. При синтезі дзвінких звуків на смугові фільтри надходять сигнали від генератора основного тону, а при синтезі глухих звуків – від генератора шуму ГШ. Сигнали тон-шум, які надходять від ВТШ по каналу зв′язку, керують ЕК.

На амплітудні модулятори (АМ1…АМ*N*) подають сигнали збудження, що пройшли крізь смугові фільтри, та інформацію про середню інтенсивність сигналу в частотних смугах, яка надходить з декодера синтезатора. Між амплітудними модуляторами і суматором ∑ включені смугові фільтри для усунення небажаних продуктів модуляції. Ці смугові фільтри на рис. 5.1 не показані. Сигнал мови на виході суматора синтезатора тим краще відображає натуральність звучання вхідного сигналу вокодера, чим більше число спектральних каналів. Як, правило, число частотних каналів вокодерів вибирають в межах 7…20. Смугові вокодери працюють зі швидкостями передач 1200…9600 біт/с. При цьому близько 600 біт/с необхідно для передачі сигналів ОТ та тон-шум [6].

Технічні вимоги до окремих вузлів смугових вокодерів можна знизити, якщо забезпечити компандування динамічного діапазону. В смугових вокодерах, як правило, використовують складовий компандер, хоча можуть застосовуватись компандери за середнім рівнем сигналу і звукові. Динамічний діапазон сигналів стискують до 18…20 дБ.

З метою забезпечення високої розбірливості мови смуги пропускання смугових фільтрів вибирають близькими до ширини смуг рівної розбірливості. При цьому може статися, що в смугу одного фільтра потрапить декілька гармонік основного тону мови, що призведе до зниження якості мовного сигналу. Застосування однакових вузькосмугових фільтрів ускладнює апаратуру.

Використання смугових фільтрів з різними смугами пропускання призводить до того, що спектральні складові сигналу проходять через приймально–передавальний тракт вокодера за різні відрізки часу, тобто смуговий вокодер вносить суттєві фазові спотворення, які оцінюються груповим часом запізнення.

Згідно з рекомендаціями МККТТ спотворення нормують для граничних частот діапазону. Фазові спотворення являють собою різницю за часом передачі сигналів з середньою частотою $\omega_{сер}$ і граничними частотами діапазону

$$\Delta t_{н} = \left|\frac{d\varphi}{d\omega}\right|_{\omega \to \omega_{н}} - \left|\frac{d\varphi}{d\omega}\right|_{\omega \to \omega_{сер}}, \qquad (5.2)$$

$$\Delta t_{в} = \left|\frac{d\varphi}{d\omega}\right|_{\omega \to \omega_{в}} - \left|\frac{d\varphi}{d\omega}\right|_{\omega \to \omega_{сер}}, \qquad (5.3)$$

де $\Delta t_{н}$, $\Delta t_{в}$ – груповий час запізнення в області нижніх $\omega_{н}$ і верхніх $\omega_{в}$ частот.

Фазові спотворення помітно спотворюють форму сигналу. Невеликі фазові спотворення не сприймаються слухом, а при великих фазових спотвореннях змінюється тембр сигналу, в кінці складу відчуваються переддихання, погіршується розбірливість мови. Дефекти мови особливо відчутні на слух, якщо груповий час запізнення наближається до тривалості складу.

Враховуючи викладене, в області частот нижче 1000 Гц використовують декілька фільтрів з рівними смугами пропускання, і тільки в області більш високих частот смуги фільтрів розширюють відповідно зі смугами рівної розбірливості. При виборі фільтрів з великою крутістю згасання поза смугою пропускання підвищується точність вимірювання характеристик спектрів сигналу, але зростає час перехідних процесів у смугових фільтрах, що спотворює швидкі зміни спектральних складових. На слух це сприймається як введення реверберації в мовний сигнал. Тому в смугових вокодерах використовуються фільтри третього – шостого порядків, які мають характеристики без сплесків. На середній частоті суміжного фільтра згасання даного фільтра вибирають близько 16… 20 дБ, а в точці перетину характеристик суміжних фільтрів –3 дБ. Фазова характеристика в смузі пропускання фільтра повинна бути лінійною з нерівномірністю ±5%, а нерівномірність характеристики групового часу запізнення всієї гребінки фільтрів повинна бути не більша 1 мс.

Фільтри нижніх частот вокодера повинні забезпечити передачу часової обвідної мовного сигналу при достатньому пригніченні гармонік основного тону. В результаті статистичних вимірювань встановлено, що 99% енергії часової обвідної знаходиться в смузі 25 Гц, тому смугу пропускання фільтрів нижніх частот вибирають близько 25 Гц. Характеристики ФНЧ повинні бути монотонними. Часто застосовують інтегратори у вигляді RC-фільтрів, які мають сталі часу близько 30 мс.

Тривалість дзвінких звуків мови більша, ніж глухих. RC-фільтр дозволяє легко виконувати інтегрування з різними сталими часу, якими можна керувати за допомогою сигналів тон-шум. На дзвінких звуках зменшують сталу часу RC-фільтра при зменшенні періодів ОТ мови.

З метою спрощення схем фільтрів і модуляторів синтезатора вирівнюють спектральні складові основного тону, які надходять через смугові фільтри на модулятори. Це призводить також до покращення звучання синтезованої мови.

Для того, щоб не скорочувались паузи мови та не спотворювались фронти сигналів, відліки параметрів у кожній з частотних смуг аналізатора беруться одночасно і подаються на модулятори синтезатора. Передача параметрів по каналу зв′язку здійснюється по черзі шляхом підключення фільтрів нижніх частот за допомогою комутатора до кодувального пристрою. Забезпечення такої послідовності операцій під час аналізу, синтезу та передачі мови можливо тільки тоді, коли є запам′ятовувальний пристрій в кодері та декодері.

Ефективність вокодерних перетворювачів оцінюють відношенням розбірливості синтезованої мови до розбірливості натуральної мови за одних і тих самих умов передачі та прийому.

$$K_{bn} = \frac{A_{bn}}{z\, A_n}, \tag{5.4}$$

де K_{bn} – коефіцієнт ефективності вокодерного перетворення; A_{bn}, A_n – формантна розбірливість синтезованої і натуральної мови в n-й смузі частот; z – коефіцієнт експериментальної практики артикуляційної бригади.

Дослідним шляхом встановлено, що артикуляційні бригади отримують стійкі результати після трьох днів тренувань до сприйняття натуральної мови і після шести днів тренувань до сприйняття синтезованої мови, тобто навики до сприйняття синтезованої мови збільшуються вдвічі повільніше. Для тренованої бригади величина z повинна бути не меншою, ніж 0,97 [6].

Тренування проводяться на тренувально-контрольному тракті з відомою величиною формантної розбірливості A_0. Коефіцієнт експериментальної практики артикуляційної бригади

$$z = \frac{A_l}{A_0}, \tag{5.5}$$

де A_l – формантна розбірливість, отримана бригадою під час даного експерименту.

Якщо вважати, що коефіцієнт експериментальної практики тренованої артикуляційної бригади близький до одиниці, то коефіцієнт ефективності воко-

дерного перетворення за оптимальних умов передачі і прийому має фізичний зміст коефіцієнта сприйняття перетвореної мови,

$$A_{bn} = K_{bn}\ A_n. \tag{5.6}$$

Наявність шумів у кожній смузі рівної розбірливості може бути врахована за допомогою коефіцієнта розбірливості формант ω_n. Тоді

$$A_{bn} = K_{bn}\ A_n\ \omega_n. \tag{5.7}$$

Якщо в смуговому вокодері використані 20 смуг рівної розбірливості, то формантна розбірливість синтезованої мови A_b буде

$$A_b = 0{,}05\sum_{n=1}^{20} K_{bn}\omega_n, \tag{5.8}$$

тому що формантна розбірливість у цьому випадку в кожній смузі максимальна, тобто в кожній смузі $A_n = 0{,}05$.

Якщо передача всіх смуг рівної розбірливості здійснюється однаково, то

$$A_b = 0{,}05\,K_{bn}\sum_{n=1}^{20}\omega_n. \tag{5.9}$$

За формулою (5.9) можна визначити максимально можливу розбірливість мови вокодера, коли число смуг аналізу і синтезу мови дорівнює 20, смуги фільтрів відповідають смугам рівної розбірливості (див. табл. 3.2), а аналіз і синтез сигналу здійснюється в смузі частот звукових сигналів до 7000 Гц.

Якщо вокодер має 20 частотних каналів, смуги яких відповідають смугам рівної розбірливості, то коефіцієнт вокодерного перетворення дорівнює 0,9. Формантна розбірливість мови в усій смузі частот вокодера теж дорівнює 0,9. При зменшенні смуг частотних каналів знижується величина K_{bn}. Так, наприклад, при звуженні смуг в шість разів величина коефіцієнта ефективності вокодерного перетворення зменшується до 0,2.

Оцінку спотворень, які вносять смугові вокодери, проводять за допомогою діагностичних таблиць. Така таблиця для ознаки «дзвінкий – глухий» налічує 400 складів, у половини яких змінюється початковий приголосний складу, а в іншій половині таблиці – кінцевий приголосний звук складу. Таблиця містить однакову кількість дзвінких та глухих приголосних звуків. Склади в таблиці мають такий вигляд: бис-пис, дим-тим, криз-крис і т.д.

Часові обвідні частотних каналів смугового вокодера корельовані між собою, тому що частотні характеристики згасання смугових фільтрів мають область перекривання, до того ж, кореляція обумовлена і характером збудження звукових сигналів мовоутворюючим трактом. Декореляція сигналів дозволяє зменшити необхідну пропускну здатність каналу зв′язку і підвищити стиснення сигналів смуговими вокодерами.

Зменшити кореляцію обвідних частотних каналів дозволяє динамічне кодування, яке полягає в тому, що за допомогою ІКМ кодують обвідну того ча-

стотного каналу, в якому в даний момент часу спостерігається максимальна амплітуда $U_{макс}$. Значення амплітуд обвідних U_i в інших каналах кодують у вигляді різниці сигналів $U_{макс} - U_i$, тобто в даному випадку використовують ІКМ і диференційне кодування. Таке кодування дозволяє обійтись меншою значністю коду при передачі різницевих сигналів та, відповідно, зменшити необхідну пропускну здатність каналу зв′язку.

Динамічне кодування дуже ефективне при кодуванні сигналів за окремими формантними областями, у яких за $U_{макс}$ беруть максимальні значення обвідних у кожній з трьох формантних областей. В інших каналах кодують тільки різниці амплітуд обвідних формантних областей.

Декореляцію сигналів смугового вокодера можна здійснити, використовуючи ортогональні перетворення. Найвідомішим є перетворення Уолша–Адамара, яке можна виконати тільки з використанням додавання та віднімання. Просто реалізується і перетворення Радемахера і Хаара. Застосування ортогональних перетворень часових обвідних частотних каналів смугового вокодера дозволяє зменшити потрібну пропускну здатність каналу зв′язку більше ніж у два рази. Якщо брати відліки параметрів частотних каналів зі змінною частотою, числове значення якої залежить від параметрів похідної спектральної обвідної, то можна зменшити швидкість передачі без помітного погіршення якості мови. Використання змінної частоти дискретизації дозволяє отримати прийнятну якість мови навіть при передачі сигналів смугового вокодера зі швидкістю 1200 біт/с.

Використовують цифрові вокодери, в яких виділення параметрів обвідних сигналів у частотних смугах здійснюється в цифровій формі. Для визначення амплітуд спектральних складових мовних сигналів застосовують дискретне перетворення Фур′є.

5.4. Формантні методи

Принцип перетворення сигналів формантного вокодера аналогічний принципу мовоутворення. Мовний тракт людини являє собою набір резонаторів, в яких змінюється добротність і резонансні частоти при вимові звуків. Параметрами резонаторів керують сигнали, які надходять з мовного центру мозку людини. При передачі сигналів в аналізаторі формантного вокодера виділяють керуючі параметри, які впливають на резонансні контури синтезатора, дозволяючи при прийомі відтворювати обвідні спектра звуків мови [6].

Форманти характеризуються амплітудою, частотою і шириною спектра звуку в тій області частот, де знаходиться дана форманта. Основою формантного уявлення є, по-перше, те, що інформація про розбірливість мови міститься в спектрально-часових обвідних мовних сигналів, а, по-друге, параметри звуків мови, які є в амплітуді, частоті і ширині спектра в області формант, незалежні. Кожному звуку відповідає цілком певна комбінація частот і часова обвідна спектра, що дозволяє розпізнавати звуки мови.

У формантних вокодерах для передачі звуків мови використовують не більше трьох-чотирьох формант. Певну роль в передачі відіграють і антиформанти, які представляють ті ділянки обвідної спектра, де є мінімум спектральних складових звуків мови. Антиформанти, в основному, характеризують носові і глухі звуки мови. Насиченість формантами ділянок частотного діапазону різна. Найбільша ймовірність появи формант спостерігається в області частот до 500 Гц. При зростанні частоти ймовірність появи формант повільно знижується. Ділянки частотного діапазону, де з′являються конкретні форманти звуків мови, значно перекриваються, що ускладнює виділення формант з мовних сигналів. При побудові вокодерів частотний діапазон мовних сигналів ділять на декілька областей з таким розрахунком, щоб в кожній частотній області була одна форманта.

Керування резонансними контурами синтезатора дозволяє точно відтворити спектральну обвідну звука мови, якщо одночасно змінювати амплітуди коливань, добротність і резонансні частоти контурів. В деяких вокодерах обмежуються тільки передачею амплітуд і частот трьох формант, що трохи погіршує якість звучання мови. Для синтезу мови важливі такі динамічні характеристики формант, як знак і швидкість зміни частоти форманти, знак і швидкість зміни її амплітуди.

Діапазон першої форманти дзвінких звуків обмежується частотами 150…900 Гц, другої – 550…2800 Гц, третьої – 1500…3500 Гц. Форманти глухих звуків мають менш визначені значення через нерівномірності спектральних обвідних. Перша форманта глухих звуків розміщена в області частот 1000… 4500 Гц, а друга форманта – 2500…10000 Гц. Антиформанта глухих звуків знаходиться між першою та другою формантами, що відповідає області частот 1500…4000 Гц [6].

Умовний динамічний діапазон сигналів першої форманти неспотвореної мови з ймовірністю 0,9 дорівнює 28 дБ, другої – 31 дБ, третьої – 25 дБ. Перед подачею сигналів на вхід вокодера часто проводять корекцію спектра з крутістю 6 дБ/октаву. Це призводить до вирівнювання динамічних діапазонів формант. Після амплітудно-частотної корекції динамічний діапазон сигналів усіх формант складає близько 25 дБ.

Діапазон змін частот формант при вимові окремих фраз мови відносно невеликий і складає 6,7…8,7 Гц для першої форманти та 7,3…8,4 Гц для другої форманти. Для передачі таких змін частот стала часу вокодера при передачі першої форманти повинна бути близько 30 мс, а при передачі другої форманти – 50 мс. Коефіцієнти кореляції амплітуд першої та другої формант в середньому дорівнюють 0,77, а третьої та четвертої формант – 0,96 [6].

Ширина смуги частот голосних звуків на рівні 3 дБ, в якій знаходиться перша форманта, складає близько 50 Гц, для другої форманти ця смуга ширша і дорівнює 72 Гц, для третьої форманти – 125 Гц. Ці смуги частот за своєю суттю є викидами обвідної спектра, які спостерігаються в області розміщення формант. Викиди обвідних спектра приголосних звуків менш виражені, тому що спектр обвідної приголосних звуків нерівномірний, а максимуми спектра не

завжди співпадають з формантною частотою, а за наявності декількох максимумів вибір формантної частоти є довільним.

Відомі два методи аналізу формант. За першого методу аналізу мовний сигнал ділять на формантні області і потім подають на блок, де визначаються амплітуди і частоти формант. Другий метод передбачає розподілення мовного сигналу вузькосмуговими фільтрами, пошук максимумів спектра, визначення амплітуд і частот сигналів з максимальними амплітудами. Також у формантних вокодерах застосовуються методи, які є модифікацією розглянутих двох методів.

При розпізнаванні звуків мови найбільшу роль відіграє величина частоти форманти, а амплітуда форманти і ширина смуги викиду спектра формантної області відіграють меншу роль. Тому в вокодерах найбільшу увагу приділяють точному визначенню частоти форманти.

Найпростіше вимірювання формантної частоти виконується за допомогою ρ-метра (рометра). Принцип роботи рометра такий, як і частотоміра. Сигнал в рометрі підлягає граничному амплітудному обмеженню, потім проводиться диференціювання обмеженого сигналу. Отримані сигнали подають на пристрій, який розширює тривалість диференційованих імпульсів. Після випрямлення та усереднення розширених імпульсів отримують напругу, величина якої пропорційна частоті вимірюваної форманти.

Структурна схема вимірювання частоти формант наведена на рис. 5.2.

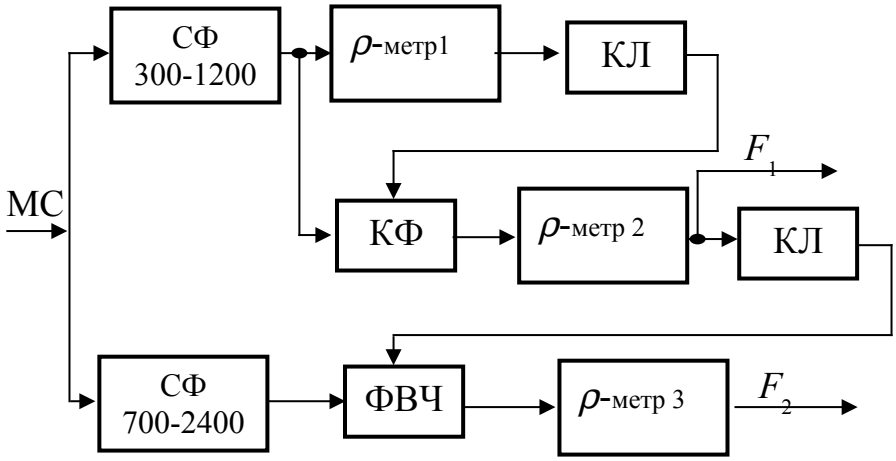


Рисунок 5.2 – Структурна схема вимірювання частоти формант за допомогою рометра:
СФ – смуговий фільтр; КФ – керуючий фільтр; ФВЧ – фільтр високих частот; КЛ – керуючий ланцюжок; F_1 та F_2 –частоти першої та другої формант

Мовний сигнал (МС) подається на два смугових фільтри (рис. 5.2). Смуги пропускання частот смугових фільтрів СФ перекриваються. Перший з фільтрів пропускає мовні сигнали з частотами 300…1200 Гц, другий фільтр – 700…2400 Гц. За допомогою першого рометра проводиться грубе визначення частоти F_1 першої форманти в широкому діапазону частот формантної області. До виходу першого рометра підключений керуючий ланцюжок (КЛ), величина вихідної напруги якого пропорційна частоті першої форманти, визна-

ченої приблизно. За допомогою напруги керуючого ланцюжка проводиться підстроювання керуючого фільтра (КФ), який пропускає вузьку смугу частот формантної області. Точне вимірювання частоти F_1 першої форманти виконується другим рометром, причому вимірювання проводиться у вузькій смузі частот.

Таке підстроювання частоти необхідне тільки для першої форманти. Частота F_2 другої форманти визначається третім рометром. Фільтр високих частот (ФВЧ) настроюється на нижню границю другої форманти за напругою, яка надходить через керуючий ланцюжок від другого рометра. Частота другої форманти визначається після того, як була визначена частота першої форманти. За необхідності вимірювання частоти третьої форманти включають додатковий фільтр зі смугою пропускання 1100…3500 Гц, а вимірювання частоти третьої форманти виконують після того, як виміряні послідовно частоти першої та другої формант.

Помилки у визначенні формантних частот за допомогою рометра тим менші, чим менший відносний інтервал частот між аналізованими складовими. Тому формантні частоти вимірюють після зміщення звукових сигналів в область високих частот (20…100 кГц). Такий метод визначення формантних частот називають методом високочастотного (ВЧ) рометра. Точність визначення частот ВЧ рометром в найгіршому випадку, коли формантні частоти знаходяться на краях формантних діапазонів, вища ніж при визначенні низькочастотним (НЧ) рометром в найкращому випадку.

Частоти формант можна вимірювати за допомогою дискримінаційного методу, за якого сигнал спочатку подають на амплітудний обмежувач, а потім – на частотний детектор. Напруга на виході частотного детектора буде пропорційна частоті сигналу.

Розроблений метод визначення середньозваженої частоти спектра, який використовують при аналізі приголосних звуків. При цьому спектр сигналів переміщують в область частот вище 20 кГц. Якщо вибрати часовий зсув сигналів τ при вимірюванні коефіцієнта кореляції так, щоб кожна складова спектра задовольняла умові $\cos \omega\tau \approx \omega\tau - 1{,}5\pi$, то буде лінійна залежність між коефіцієнтом кореляції та середньозваженою частотою спектра. Вимірявши коефіцієнт кореляції при зсуві сигналу на величину τ, можна визначити середньозважену частоту спектра приголосного звуку.

Один з методів другої групи аналізу формант полягає у тому, що за допомогою комплекту вузькосмугових фільтрів ділять мовний сигнал на велику кількість смуг. Виходи фільтрів комутують зі швидкістю переключення до 100 разів за секунду з метою пошуку спектральних максимумів у діапазоні частот. Знайдені таким чином максимуми спектральної обвідної мовного сигналу і будуть формантами.

Вузькосмугові фільтри можуть бути згруповані за трьома формантними областями, а саме: 1) 150…800 Гц; 2) 800…2280 Гц; 3) 2280…7000 Гц. При переключенні фільтрів в кожній області знаходять форманти за відповідними максимумами спектра.

Визначення формант за допомогою фільтрів може бути проведено з точністю до ширини смуги фільтра. Для підвищення точності визначення ча-

стот формант використовують інтерполяційну схему, яка дозволяє точніше знаходити місцеположення максимальної амплітуди спектральної обвідної мовних сигналів. Інтерполяційна схема дозволяє порівнювати напруги в сусідніх каналах і в усіх інших каналах, де напруга відрізняється не більше ніж на 6 дБ. Достоїнство схеми не тільки в більш точному визначенні формантної частоти, а також в можливості визначення амплітуди форманти.

Для вимірювання частот формант можна використовувати кореляційний метод. Лінія затримки корелометра в даному випадку повинна розраховуватись на більш високий частотний діапазон, ніж при вимірюванні корелометром основного тону мови. Достатньо точно можна виміряти частоти формант і за допомогою кепстрального аналізу, використовуваного при визначенні основного тону мовних сигналів. Кепстр має максимуми на частотах формант і мінімуми при антиформантах.

Дуже точно можна визначити частоти формант, якщо ввести в аналізатор вокодера формантний синтезатор з перестроюванними контурами, якими керують параметри формант, виділені аналізатором. Компаратори порівнюють сигнали аналізатора та синтезатора, а помилки порівняння використовують для підстроювання формантних частот, що дозволяє точно визначити частоти формант.

Розрізнюють формантні синтезатори з послідовним та комбінованим (послідовно-паралельним) з′єднанням формантних контурів. На рис. 5.3 показані два варіанти побудови синтезаторів з комбінованим з′єднанням контурів. Додаткові керовані параметри другого варіанта на схемі показані пунктиром.

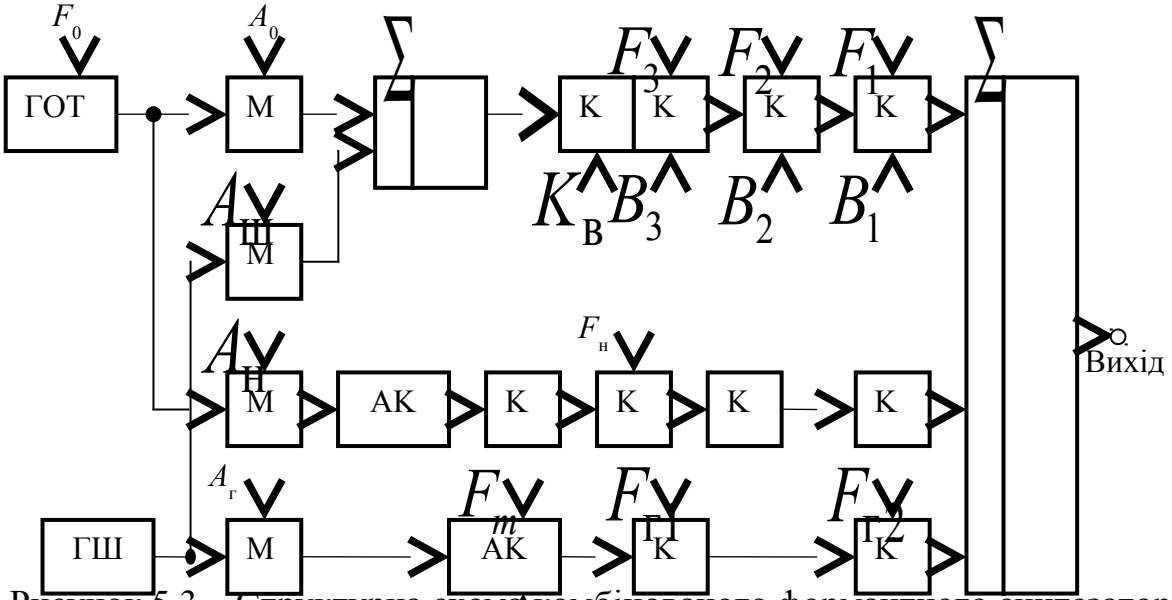


Рисунок 5.3 – Структурна схема комбінованого формантного синтезатора:

ГОТ – генератор основного тону; ГШ – генератор шуму; М – модулятор; К – формантні контури; ∑ – суматор; АК – антиформантні контури; F_0 та A_0 – частота та амплітуда основного тону; $A_ш$ – амплітуда шумових сигналів; $F_н$ та $A_н$ – частота та амплітуда носових звуків; $A_г$ – амплітуда глухих звуків; F_1, F_2, F_3 – частоти першої, другої і третьої формант дзвінких звуків; F_m – антиформанта частота глухих звуків; $F_{г1}$ і $F_{г2}$ – перша і друга частоти формант глухих звуків; $K_в$ – корекція спектра для формант вище третьої; B_1, B_2, B_3 – ширина смуг формант дзвінких звуків; B_m – ширина смуги антиформанти глухих звуків

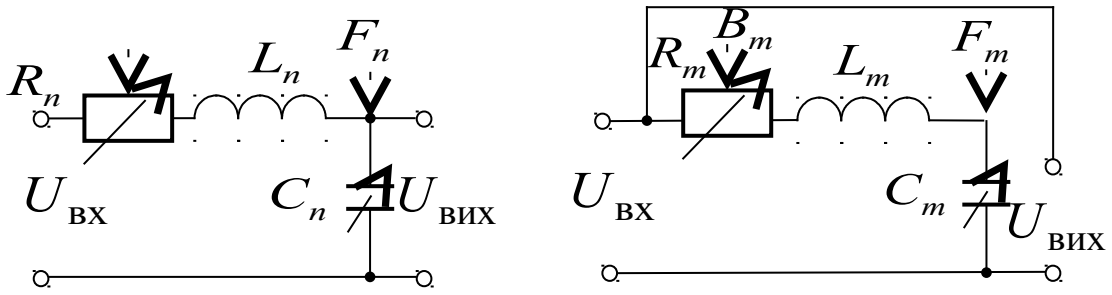
Перший варіант формантного синтезатора розрахований на пропускну здатність каналу зв′язку 1200 біт/с і забезпечує задовільну якість звучання. Дру-

гий варіант синтезатора забезпечує відмінне звучання і вимагає пропускної здатності каналу зв′язку 2400 біт/с.

Структурна схема синтезатора являє собою три ланцюжки, з′єднані паралельно. Кожен з ланцюжків містить послідовно підключені контури, які дозволяють синтезувати дзвінкі, носові і глухі звуки. Спроби створення формантного синтезатора з одним послідовним ланцюжком привели до негативних результатів через складності керування контурами такого синтезатора.

Дзвінкі звуки синтезуються за допомогою сигналів основного тону та трьох формантних частот. Кількість контурів у ланцюжку можу бути від трьох до п′яти, хоча керування здійснюється тільки трьома контурами.

Схема формантного контура показана на рис. 5.4, а.



а) б)

Рисунок 5.4 – Схема контурів:
а) – формантний, б) – антиформантний

Паралельний контур, складений з елементів R_n, C_n, L_n підстроюється на резонансну частоту, яка дорівнює формантній частоті, за допомогою ємності C_n. Підстроєння здійснюється за допомогою параметра частоти форманти F_n, який виділили аналізатором та передали по каналу зв′язку. Зміна ширини смуги форманти проводиться шляхом зміни добротності контура за допомогою активного опору R_n. Розміром опору R_n керує параметр B_n ширини смуги форманти, виділений при передачі і отриманий по каналу зв′язку. На частоті резонансу вихідна напруга формантного контура буде максимальною.

При синтезі дзвінких приголосних на вхід ланцюжка подається невелика шумова складова від генератора шуму, це забезпечує натуральність звучання мови. Керування шумовою складовою здійснюється за допомогою параметра $A_ш$.

Спектр глухих звуків мови має форманти та антиформанти. В області формант спектральні складові максимальні, а в області антиформант – мінімальні. Ланцюжок для синтезу глухих звуків містить, як правило, два формантних і один або два антиформантних контури. Формантними контурами глухих звуків керують параметри частоти F_m і ширина смуги B_m антиформанти.

Антиформантний контур являє собою послідовне з′єднання елементів R_m, C_m, L_m (див. рис. 5.4, б), яке забезпечує отримання мінімальної вихідної напруги на частоті антиформанти. На вхід ланцюжка синтезатора глухих звуків подають

шумовий сигнал, величина якого залежить від параметра $A_г$, який визначає рівень глухих звуків. Частоти першої та другої формант глухих звуків залежать від параметрів $F_{г1}$ і $F_{г2}$. В схемі, показаній на рис. 5.3, керування смугами формант глухих звуків не проводиться, а змінюють тільки частоти формант, тому що смуги формант глухих звуків дуже широкі.

Ланцюжок синтезу носових звуків складається з декількох формантних і одного або двох антиформантних контурів. Частоти та смуги цих контурів вибираються іншими, ніж при синтезі дзвінких і глухих звуків. Частоти антиформант носових звуків можна не переміщувати по частотному діапазону, а з усіх формант змінюють тільки частоту $F_н$ однієї форманти. Рівень носових звуків, подаваних на вхід ланцюжка, визначається параметром $A_н$.

У першому варіанті синтезатора використовують такі параметри: частоти трьох формант звуків F_1, F_2, F_3; частота F_0 і амплітуда A_0 основного тону; амплітуда $A_г$ глухих звуків; амплітуда $A_н$ носових звуків; амплітуда $A_ш$ дзвінких шумових звуків; частота $F_{г1}$ першої форманти глухих звуків; різниця $F_{г2}– F_{г1}$ між частотами другої і першої формант глухих звуків; відношення $F_m/F_{г1}$ антиформантної і першої формантної частот глухих звуків; сигнал тон-шум. У ланцюжку дзвінких звуків містяться контури для четвертої і п′ятої формант з постійним настроюванням і коректор спектра для формант вище третьої. Дані про параметри першого варіанту формантного вокодера наведені в табл. 5.1. Параметр тон-шум можна не передавати, тому що при прийомі його можна відновити за моментами пропадання або появи ОТ.

Таблиця 5.1 – Параметри першого варіанту вокодера з комбінованим з′єднанням формантних контурів

Параметр	Кількість		Межі вимірювань, Гц, та інтервали
	біт	рівнів	між рівнями, дБ
F_0	4	16	60…340, через 1/16 октави
F_1	4	16	150…900, через 50 Гц
F_2	4	16	550…2800, через 150 Гц
F_3	3	8	1550…4000 Гц, через 350 Гц
$F_{г1}$	2	4	3000…6000, через 1000 Гц
$F_{г2}–F_{г1}$	1	2	2000 або 3000 Гц
$F_m/F_{г1}$	1	2	0,5 або 1,41
A_0	3	8	5…25, через 5 дБ і – ∞
$A_г$	3	8	5…25, через 5 дБ, і – ∞
$A_ш$	2	4	–5, –10, –20 дБ або –∞
$A_н$	2	4	–5, –10, –20 дБ або – ∞
Тон-шум	1	2	0 або ∞

Як видно з таблиці, для передачі параметрів вокодера треба 30 біт, що при 40 відліках за секунду вимагає 1200 біт/с пропускної здатності каналу зв′язку.

Другий варіант синтезатора забезпечує високу якість мови, кількість передаваних параметрів більша, а для передачі необхідно мати канал з пропускною здатністю 2400 біт/с. Параметри синтезатора та необхідні пояснення наведені в табл. 5.2.

Таблиця 5.2 – Параметри другого варіанта вокодера з комбінованим з′єднанням формантних контурів

Параметр	Межі змін	Точність	Примітки
F_0	50…300 Гц	3%	Частота ОТ
F_1	200…1200 Гц	3%	Формантні частоти дзвінких,
F_2	500…3100 Гц	3%	крім носових
F_3	1000…6200 Гц	3%	
A_0	0…32 дБ	1 дБ	Рівень дзвінких, крім носових
$A_г$	0…28 дБ	4 дБ	Рівень глухих
$A_ш$	0…28 дБ	4 дБ	Рівень шумів дзвінких приголосних
$A_н$	0…24 дБ	8 дБ	Рівень носових
$F_н$	200…1100 Гц	12%	Формантні частоти носових
$K_в$	2500…6200 Гц	12%	Корекція вищих формант
F_m	1000…6200 Гц	3%	Антиформантна частота глухих
$F_{г1}$	1000…6200 Гц	3%	Формантні частоти
$F_{г2}$	2500…9800 Гц	3%	глухих
B_1		100 Гц	Ширина формант
B_2		100 Гц	дзвінких, крім
B_3		200 Гц	носових

Структурна схема формантного вокодера з паралельним аналізом і синтезом показана на рис. 5.5.

При паралельному синтезі звуків мови немає окремих ланцюжків для синтезу різних груп звуків. Вузькосмугові фільтри в аналізаторі вокодера згруповані за трьома СФ1, СФ2, СФ3 формантними областями, в кожній з яких міститься від 6 до 11 смугових фільтрів. У процесі аналізу звука мови в кожній з груп смугових фільтрів за допомогою амплітудних детекторів АД1, АД2, АД3 знаходять максимуми спектрів А1, А2, А3, які відповідають амплітудам першої, другої та третьої формант. Частоти F_1, F_2, F_3 відповідних формант визначають за допомогою детекторів формантних частот ЧД1, ЧД2, ЧД3.

В аналізаторі виділяють частоту основного тону F_0 і амплітуду шумових звуків $A_ш$. Для цієї мети в аналізаторі є виділяч основного тону (ВОТ) і виділяч тон-шум ВТШ.

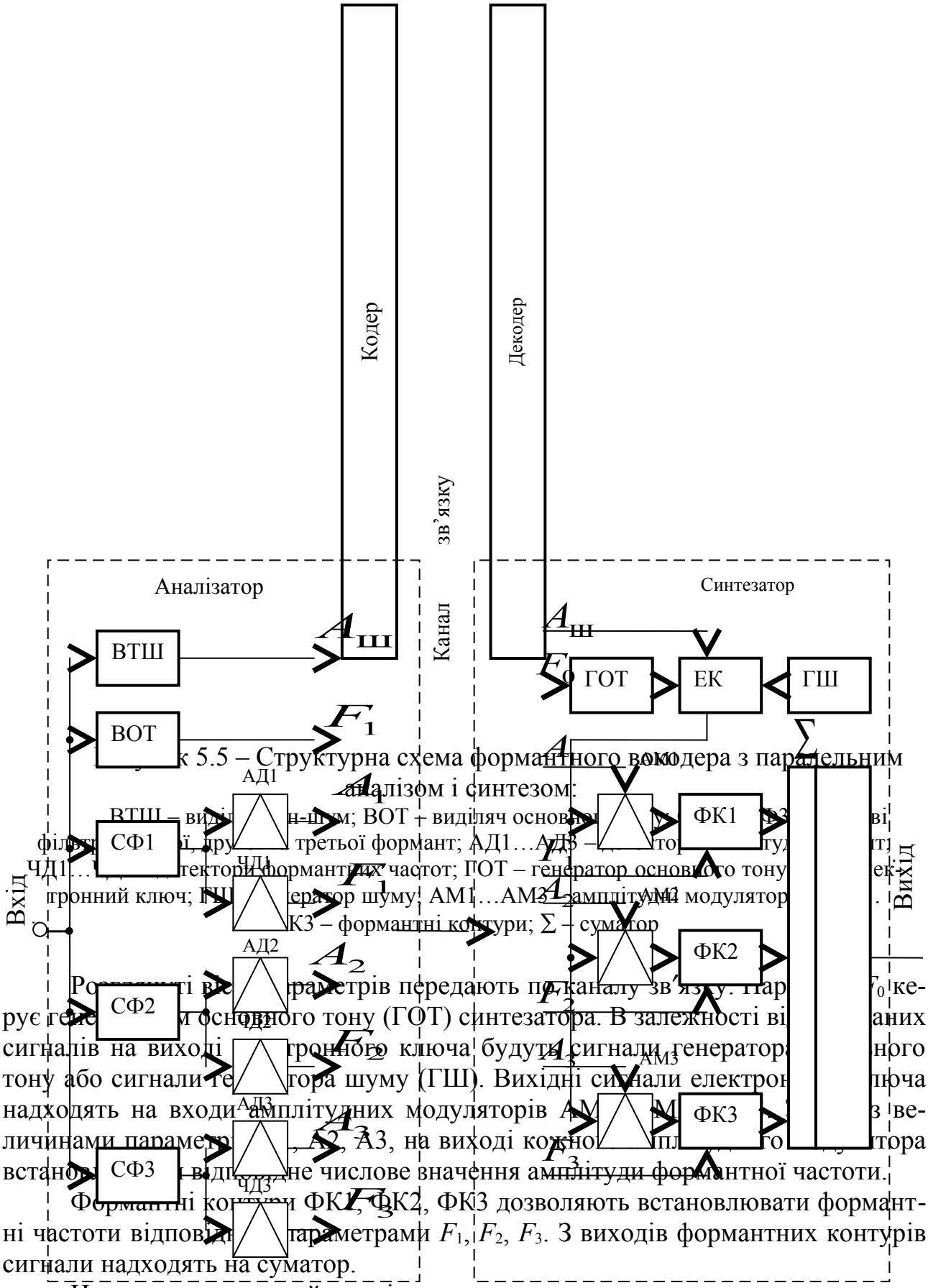


Рисунок 5.5 – Структурна схема формантного вокодера з паралельним аналізом і синтезом:

ВТШ – виділяч шум-шум; ВОТ – виділяч основного тону; СФ1…СФ3 – смугові фільтри 1-ої, другої, третьої формант; АД1…АД3 – детектори амплітуд; ЧД1…ЧД3 – детектори формантних частот; ГОТ – генератор основного тону; ЕК – електронний ключ; ГШ – генератор шуму; АМ1…АМ3 – амплітудні модулятори; ФК1…ФК3 – формантні контури; Σ – суматор

Розглянуті вище параметрів передають по каналу зв'язку. Параметр F_0 керує генератором основного тону (ГОТ) синтезатора. В залежності від керуючих сигналів на виході електронного ключа будуть сигнали генератора основного тону або сигнали генератора шуму (ГШ). Вихідні сигнали електронного ключа надходять на входи амплітудних модуляторів АМ1…АМ3, на інші входи з величинами параметрів A_1, A_2, A_3, на виході кожного амплітудного модулятора встановлюється відповідне числове значення амплітуди формантної частоти.

Формантні контури ФК1, ФК2, ФК3 дозволяють встановлювати формантні частоти відповідними параметрами F_1, F_2, F_3. З виходів формантних контурів сигнали надходять на суматор.

Наведена схема є найпростішою, але вона дозволяє розглянути паралельний синтез звуків. Комбінований метод синтезу звуків складніший, але якість

синтезованої мови краща. Тому паралельний синтез використовують при передачі службової звукової інформації. Якщо в схемі, наведеній на рис. 5.5, для передачі параметрів формант використовувати по 3 біт, а для передачі частоти основного тону та сигналу тон-шум – по 6 біт, то при 40 відліках за секунду для організації телефонного каналу треба $(6 \times 2 + 6 \times 3) \times 40 = 1200$ біт/с.

Для зменшення габаритів апаратури та спрощення схемних рішень аналіз і синтез сигналів проводять в області частот 20…26 кГц. Перенесення звукових сигналів в область високих частот виконують за допомогою однополосних амплітудних модуляторів з фільтрацією нижньої бокової смуги. Перетворення сигналів уверх за частотою і назад не змінюють часові характеристики мовних сигналів на виході синтезатора.

Формантні вокодери з цифровою обробкою сигналів створюються за тими самими принципами. Різниця лише в тому, що замість аналогових фільтрів застосовують цифрові фільтри. Сучасні швидкодіючі інтегральні схеми дозволяють проводити перестроювання кожного цифрового фільтра по черзі одним арифметичним пристроєм.

Сигнал-параметри зберігають в буферній пам’яті, з якої по черзі витягують дані для перестроювання кожного фільтра. При обробці застосовують комутатори, зсовуючі регістри, регістри затримки та накопичувачі.

При послідовному з’єднанні формантних контурів форманти формуються так само, як і в натуральному мовному тракті. Паралельне з’єднання формантних контурів призводить до того, що сумарна характеристика системи аналізу-синтезу відхилюється від ідеальної (з плоскою амплітудно-частотною та лінійною фазо-частотною характеристиками). Це призводить до того, що мова на виході паралельного синтезатора буде звучати з сильними спотвореннями, які за сприйняттям на слух схожі на реверберацію. Тому на приймальний кінець каналу зв’язку треба передавати інформацію про амплітуду та фазу форманти. Формантні вокодери, в яких передбачена передача та відновлення фази форманти, називають ***фазовими вокодерами***.

В одному з варіантів фазового вокодера в канал зв’язку передають параметри, які відповідають дійсній та уявній складовим комплексної амплітуди формант. Синтезатор фазового вокодера містить формантний контур, керований параметром частоти формант, а також – модулятори для дійсної та уявної складових амплітуди форманти. Для якісного звучання синтезованої мови необхідне керування шириною форманти, яке здійснюється введенням керованого згасання в формантний контур.

У більш ранніх розробках фазових вокодерів використовували передачу амплітуди і фази форманти. Конструктивно такі вокодери складніші, оскільки повинні містити квадратори, блоки добування квадратного кореня, блоки синусних та косинусних генераторів, помножувачі.

Формантні вокодери забезпечують приблизно вдвічі більше стиснення мовного сигналу, ніж смугові за однакових якісних показників звучання. При передачі формантної частоти її відхилення не повинно перевищувати 3%. Для пізнавання мови необхідно точно передавати максимуми спектра, навіть якщо вони не відносяться до формант. У спектрі мови даної людини можуть бути

хибні форманти, які сприяють пізнаванню мови. Якщо з мови вилучити помилкові форманти, то мова буде ідеалізованою та погіршиться пізнавання голосу абонента.

Розбірливість мови, переданої за допомогою формантних вокодерів вища, ніж при звичайному телефонному зв'язку, а натуральність звучання зберігається навіть за швидкості передачі 1200 біт/с. Натуральність звучання мови підвищується, якщо використовують корекцію в області високих та низьких частот. Мовний тракт людини складається не з трьох-чотирьох резонаторів, як це моделюється при розробці формантних вокодерів, а являє собою складну систему. Тому ряд характеристик мовного тракту не враховуються при аналізі та синтезі звуків мови. Ускладнення аналізаторів та синтезаторів формантних вокодерів у майбутньому призведе як до скорочення швидкості передачі, так і до підвищення натуральності звучання [6].

5.5. Ортогональні та спектрально-часові методи

5.5.1. Ортогональні методи

Спектральну обвідну мовного сигналу можна подати у вигляді ортогональних функцій. Система нескінченних функцій

$$\varphi_0(x),\ \varphi_1(x),\ \varphi_2(x),\ldots,\varphi_n(x),\ldots \qquad (5.10)$$

називається ***ортогональною*** на відрізку (a, b), якщо

$$\int_a^b \varphi_n(x)\varphi_m(x)dx = 0 \qquad \text{при } n \neq m. \qquad (5.11)$$

При цьому

$$\int_a^b \varphi_n^2(x)dx \neq 0, \qquad (5.12)$$

тобто жодна з функцій системи (5.10) не дорівнює нулю.

Умова (5.11) визначає попарну ортогональність функцій системи (5.10).

Вибір раціональної системи функцій залежить від мети розкладання функцій в ряд. Якщо треба точно розкласти функцію на прості ортогональні функції, то можна використати основні тригонометричні функції – синуси та косинуси. Такі функції характеризуються частотою, амплітудою та фазою. При проходженні гармонічних коливань через лінійні ланцюжки змінюється тільки амплітуда та фаза коливань, що дозволяє легко аналізувати проходження складного сигналу через такі ланцюжки.

Коли необхідно звести до мінімуму число членів ряду при представленні сигналів з заданою похибкою, використовують різні ортогональні системи функцій (поліном Чебишева, Ерміта, Лаггера, Лежандра, функції Уолша та ін.).

При виборі системи ортогональних функцій враховують можливість їх виділення при передачі та відтворення на приймальному кінці каналу зв'язку. В даний час досконально опрацьований метод представлення сумарної спектральної обвідної мовного сигналу у вигляді суми гармонічних функцій. Цей метод використовується в гармонічному вокодері, запропонованому А.А. Піроговим.

Розвиток обчислювальної техніки привів до широкого розповсюдження ортогональних функцій Уолша і Радемахера, відомих з 1922 року і надовго забутих. Використовують різні варіанти цих функцій, які відрізняються способом нумерації функцій в системі. Спосіб нумерації функцій в системі називають упорядковуванням. Застосовують упорядковування Уолша, Адамара, Пелі та ін. Позитивна якість вказаних функцій в тому, що при обробці сигналів операції множення практично виключаються і зводяться до операцій підсумовування, технічна реалізація яких простіша. Ця позитивна якість сприяла розповсюдженню функцій Уолша та Радемахера в техніці стиснення мовних сигналів, в телебаченні, радіолокації та ін.

Принцип роботи гармонічного вокодера полягає в тому, що по каналу зв'язку передають коефіцієнти розкладання в ряд спектральної обвідної сигналу, а при прийомі по цих коефіцієнтах відновлюють спектральну обвідну сигналу. Число коефіцієнтів залежить від необхідної точності відновлення обвідної.

Вхідний мовний сигнал в одному з варіантів гармонічного вокодера піддають тим самим операціям, що і в смуговому вокодері, тобто ділять мовний сигнал на смуги, детектують та проводять низькочастотну фільтрацію. В результаті таких операцій отримують спектральну обвідну мовного сигналу, яку подають на матрицю, що дозволяє перетворювати відліки спектральної обвідної в коефіцієнти ряду Фур'є. Кожні 20…25 мс беруть відліки спектральної обвідної y_k, а потім обчислюють коефіцієнти ряду Фур'є

$$a_0 = \frac{1}{2n}\sum_{k=0}^{n-1} y_k, \tag{5.13}$$

$$a_m = \frac{1}{n}\sum_{k=0}^{n-1} y_k \cos\frac{km\pi}{n}, \tag{5.14}$$

$$b_m = \frac{1}{n}\sum_{k=0}^{n-1} y_k \sin\frac{km\pi}{n}, \tag{5.15}$$

де m – індекс коефіцієнта Фур'є; a_m та b_m – коефіцієнти Фур'є; n – порядок тригонометричного ряду.

Величина n дорівнює числу смугових фільтрів, m відповідає числу формант в діапазоні частот мовного сигналу. Коефіцієнти Фур'є змінюються за часом з тією самою частотою, з якою змінюються рівні спектральної обвідної в смуговому вокодері, тобто з частотою 20…25 Гц.

Для отримання натуральності звучання мови треба передавати не менше ніж три-п'ять формант. З урахуванням передачі сталої складової α_0, кількість

передаваних коефіцієнтів гармонічного вокодера буде 7…11. В канал зв'язку також передають сигнали основного тону та сигнал тон-шум.

Оскільки в формулах (5.14), (5.15) функції синуса та косинуса не можуть бути більшими 1, а величини k, m, n, π заздалегідь відомі, то добутки y_k на вказані функції реалізуються за допомогою резисторних дільників. Точність підбору величин опорів резисторів може бути дуже високою. Коефіцієнти Фур'є потім визначаються за допомогою суматорів.

Синтез мовних коливань на приймальному кінці полягає в відновленні спектральної обвідної сигналів, які надходять від місцевих джерел коливань, якими є генератор основного тону та генератор шуму. В залежності від параметра тон-шум, отриманого по каналу зв'язку, на вхід лінії затримки синтезатора надходять сигнали основного тону або шуму.

У лінії затримки є відгалуження через рівні часові затримки τ. За приклад, на рис. 5.6 показана схема гармонічного синтезатора, в якій використана лінія затримки з сьома відгалуженнями.

Розглянемо випадок, коли на вхід лінії затримки (ЛЗ) надходять сигнали E основного тону мови з частотою ω та одиничною амплітудою, тобто

$$E = \exp(j\omega t). \tag{5.16}$$

Центральне відгалуження лінії затримки на рис. 5.6 позначене цифрою 0. Сума сигналів U_{c1} на виході суматора С, підключеного симетрично відносно центра до виводів лінії затримки з номерами 1 та –1, буде

$$U_{c1} = \exp(j\omega t)(\exp(-j\omega\tau) + \exp(j\omega\tau)) = \exp(j\omega t)(\cos\omega\tau - j\sin\omega\tau + \cos\omega\tau + j\sin\omega\tau) = \exp(j\omega t)2\cos\omega\tau. \tag{5.17}$$

Сума сигналів, взятих із наступної пари виводів з номерами 2 та –2, визначається як

$$U_{c2} = \exp(j\omega t)2\cos 2\omega\tau. \tag{5.18}$$

Якщо зменшити амплітуди сигналів вдвічі та подати їх на модулятори М, які керують амплітудами сигналів за допомогою параметрів, отриманих по каналу зв'язку, то в момент часу $t = 0$ після складання сигналів отримаємо

$$U_c = (a_0/2) + a_1\cos\omega t + a_2\cos 2\omega\tau + a_3\cos 3\omega\tau, \tag{5.19}$$

де a_0 – стала складова сигналу, яку отримують від центра лінії затримки; a_1, a_2, a_3 – коефіцієнти Фур'є.

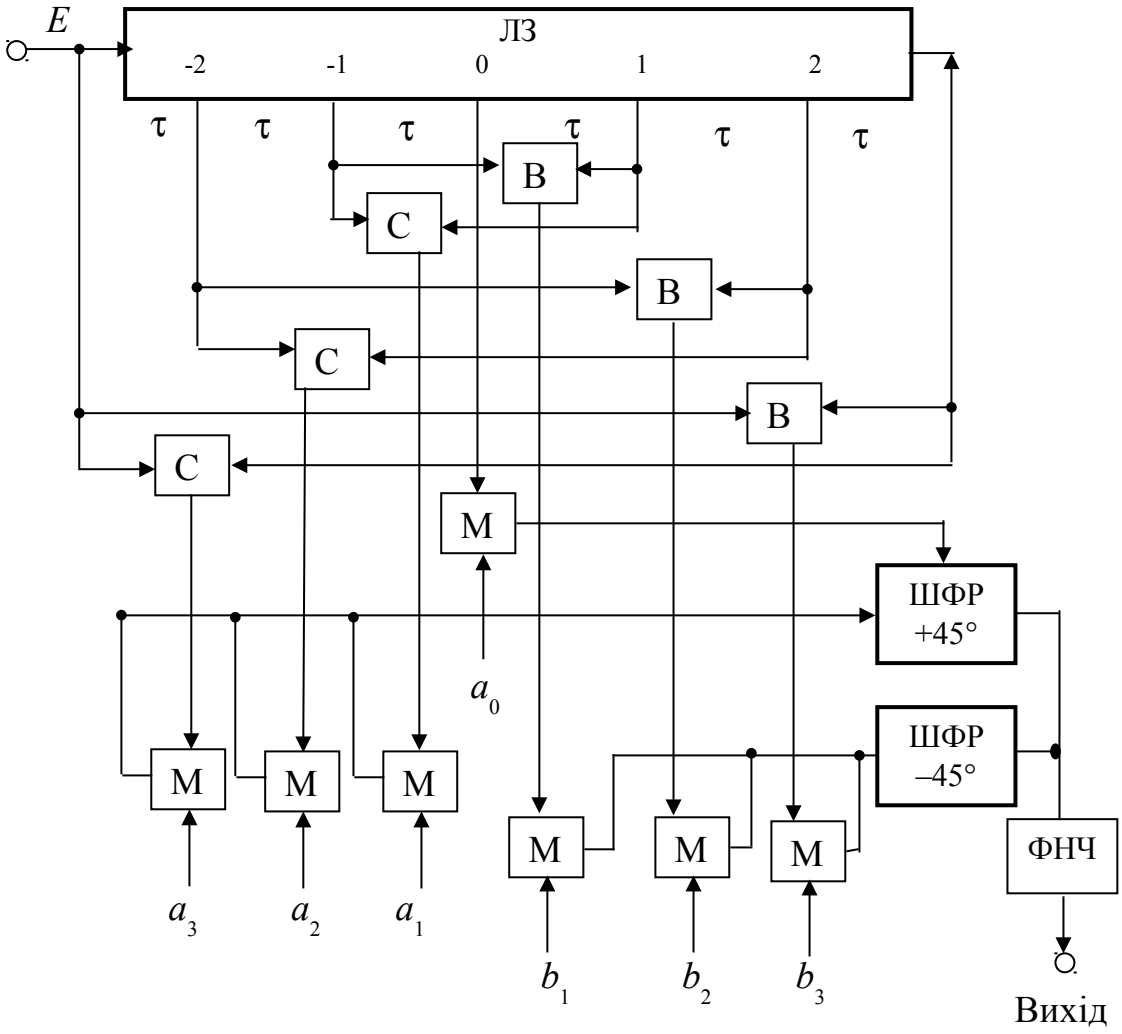


Рисунок 5.6 – Схема гармонічного синтезатора:
ЛЗ – лінія затримки; С – суматори; В – віднімачі; М – модулятори; ШФР – широкосмугові фазорозщеплювачі; ФНЧ – фільтр низької частоти

Сигнали U_{bm} на виходах віднімачів В, підключених до виводів, взятих на відстані $\pm m\tau$ від центра, можна визначити як

$$U_{bm} = \exp(j\omega t)(\exp(-jm\omega\tau) + \exp(jm\omega\tau)) = \exp(j\omega t)2j\sin m\omega\tau. \quad (5.20)$$

Склавши в момент часу $t = 0$ сигнали, які пройшли через віднімачі та модулятори, з урахуванням зменшення амплітуд сигналів вдвічі отримаємо

$$U_b = j(b_1\sin\omega\tau + b_2\sin 2\omega\tau + b_3\sin 3\omega\tau). \quad (5.21)$$

Формули (5.19) та (5.21) являють собою спектри, подані у вигляді косинусного та синусного рядів. Далі спектральні складові надходять на широкосмугові фазорозщеплювачі, які дозволяють повернути фази складових U_c та U_b на 90° відносно один одного.

На виході фільтра низьких частот ФНЧ спектр сумарного сигналу $S(\omega)$ буде мати вигляд

$$S(\omega) = \frac{a_0}{2} + \sum_{i=1}^{m} a_i \cos i\omega\tau + \sum_{i=1}^{m} b_i \sin i\omega\tau. \qquad (5.22)$$

Спектр $S(\omega)$ періодичний. Період спектра дорівнює $\Omega = 2\pi/\tau$. Фільтр низьких частот виділяє тільки складові на ділянці $0\ldots\Omega$, які відповідають діапазону мовного сигналу.

Існують вокодери, в яких використовують тільки косинусний

$$S_k(\omega) = \frac{a_0}{2} + \sum_{i=1}^{m} a_i \cos i\omega\tau \qquad (5.23)$$

або тільки синусний ряди

$$S_c(\omega) = \sum_{i=1}^{m} b_i \sin i\omega\tau. \qquad (5.24)$$

У синусному вокодері на один параметр менше, тому там відсутня стала складова. Схеми вокодерів, які використовують лише ряди (5.23) та (5.24), простіші.

Аналогові елементи довгих ліній та елементи дискретної техніки використовують як лінії затримки.

У гармонічних вокодерах можливі інтерференційні спотворення, які проявляються в посиленні або послабленні частотних складових. Підкреслимо, що певне значення в цих спотвореннях відіграють фазорозщеплювачі. Відзначається також майже однакова якість звучання синусно-косинусних, косинусних та синусних вокодерів.

Розроблений гармонічний вокодер, в якому при передачі визначаються не амплітуди, а рівні спектральної обвідної. В аналізаторі такого вокодера після кожного смугового фільтра підімкнені логарифматори. Далі сигнали обробляються за допомогою матриці, як і в звичайному гармонічному вокодері. Синтезатор вокодера побудований за послідовною схемою, що дозволяє зменшити інтерференційні спотворення.

Інколи смуги фільтрів в аналізаторі вибирають рівними ширині смуг рівної розбірливості. Оскільки ширина смуг збільшується зі зростанням середньої частоти фільтра, то на виходах фільтрів включають дільники, для того щоб при надходженні на аналізатор «білого» шуму на виходах всіх фільтрів були сигнали рівної величини. В цьому випадку можна використовувати ті самі матриці та формули для розрахунку коефіцієнтів Фур'є, що і в звичайному гармонічному вокодері. В синтезаторах таких вокодерів використовують спеціальні лінії затримки, в яких величина затримки залежить від частоти за заданим законом.

Як випливає з формул (5.16)…(5.22), лінія затримки може використовуватись як смуговий фільтр, оскільки сигнали на її виходах відповідають сигналам на виході комплекта смугових фільтрів. Важко здійснювати затримку в області низьких частот, тому частотний діапазон сигналів перед поданням на лінію затримки зміщують в область високих частот. Позитивною якістю такої фільтрації є простота конструкції, коли можна легко змінювати кількість та ширину смуг.

Є вокодер, в якому спектральна обвідна мовного сигналу виражена сумою ортогональних функцій виду sin *nx*/sin *x*. Цей вокодер по суті є варіантом гармонічного вокодера. Не виключено застосування ортогональних вокодерів, використовують й інші ортогональні функції, які згадані на початку цього розділу. Хороші результати отримані при використанні дискретних перетворень Фур'є з метою стиснення не лише мовних, але й музичних сигналів.

5.5.2. Кореляційні методи

Функція кореляції мовного сигналу може бути визначена, якщо відомий енергетичний спектр. Можна знаходити ординати спектральної обвідної за функцією кореляції та навпаки. Дискретні відліки функції кореляції $B(i\tau_0)$ аналізованого сигналу пов'язані з ординатами обвідної енергетичного спектра $G(i\omega_0)$ співвідношенням

$$G(i\omega_0) = \sum_{i=0}^{k} B(i\tau_0)\cos(i\omega_0\tau_0), \qquad (5.25)$$

де k – число відліків кореляційної функції, ω_0 – різниця частот між сусідніми ординатами спектральної обвідної, $\tau_0 = 1/2F$ (F – середня частота спектра мовного сигналу).

Функцію кореляції, визначену формулою (5.1), можна отримати за допомогою лінії затримки з відгалуженнями і помножувачів. На входи кожного з помножувачів подають початковий мовний сигнал та сигнал, затриманий на певний відрізок часу. Затриманий сигнал надходить з відгалужень лінії затримки. До помножувачів підключені інтегратори, якими можуть бути фільтри нижніх частот. Сигнали на виходах фільтрів пропорційні відлікам функції кореляції. Ці відліки як сигнал-параметри кореляційного вокодера можуть передаватись по каналу зв'язку.

У синтезаторі відліки кореляційної функції можна перетворювати в ординати енергетичного спектра за формулою (5.25). Такий перерахунок можна виконати з допомогою матриці, яка за своєю структурою схожа на матрицю перерахунку ординат спектра в коефіцієнти Фур'є за формулами (5.13)…(5.15).

Для отримання ординат спектра амплітуд обвідної мовного сигналу, треба виконати операцію добування кореню з ординат спектра потужності. Якщо в синтезаторі мови замість спектра амплітуд застосувати енергетичний спектр, то якість мови не дуже зміниться.

Один з недоліків передачі по каналу зв'язку сигнал-параметрів кореляційної функції є розширення вдвічі динамічного діапазону сигналів, які надходять

в канал зв'язку, тому що кореляційна функція пропорційна енергетичному спектру. З метою усунення цього явища сигнали, що подають на вхід лінії затримки, піддають сильній компресії, зменшуючи динамічний діапазон вдвічі. Також можна відліки кореляційної функції подавати на пристрій добування квадратного кореня. І в першому, і в другому випадках динамічний діапазон сигнал–параметрів, які надходять в канал зв'язку, відповідає динамічному діапазону відліків спектра амплітуд обвідних мовних сигналів.

В одному з варіантів кореляційного вокодера після пристрою добування квадратного кореня підключена матриця, на виході якої отримують коефіцієнти Фур'є, які визначаються за формулами (5.13)…(5.15). Далі коефіцієнти Фур'є передаються по каналу зв'язку і за ними здійснюється синтез мовних сигналів. Розроблені й інші типи вокодерів, які поєднують в собі ознаки ортогональних та кореляційних вокодерів. Окрім гармонічних функцій при аналізі та синтезі сигналів в таких вокодерах застосовують функції Лаггера, Чебишева, Лежандра, Ерміта, Уолша.

5.5.3. Гомоморфна обробка сигналів

Мовний сигнал можна розглядати як сигнал на виході лінійної системи з повільно змінними параметрами. Проблема аналізу мовних сигналів зводиться до зміни параметрів моделі мовного тракту та оцінки зміни цих параметрів за часом.

Гомоморфна обробка сигналів дозволяє визначити інформацію про сигнал збудження, що розміщена в області великого часу, від інформації про мовний тракт та форми імпульсів збудження, яка знаходиться в області малого часу кепстра. На основі кепстра можна оцінити формантні частоти, період основного тону, а також класифікувати дзвінкі та глухі звуки мови.

Розглянемо основні особливості гомоморфної обробки сигналів на конкретних прикладах. Сигнал

$$S(t) = S_1(t)\, S_2(t) \tag{5.26}$$

треба представити у вигляді суми

$$x(t) = x_1(t) + x_2(t). \tag{5.27}$$

Шуканий оператор перетворення позначимо символом D.

$$D(S(t)) = D(\mathrm{S}_1(t)\, S_2(t)) = D(S_1(t)) + D(S_2(t)). \tag{5.28}$$

Рівнянню (5.28) задовольняє логарифмічна функція, а оператор D відповідає логарифму, тобто $D(S) = \log S$, а

$$x(t) = \log(S(t)) = \log(S_1(t)\, S_2(t)) = \log(S_1(t)t) + \log(S_2(t)) = x_1(t) + x_2(t). \tag{5.29}$$

У даному випадку для спрощення припускали, що дійсні функції $S_1(t)>0$ та $S_2(t)>0$.

За спектром та формою сигнали $x_1(t)$ та $x_2(t)$ відрізняються від сигналів $S_1(t)$, $S_2(t)$, але суму сигналів, визначану формулою (5.27), можна обробляти за допомогою лінійного ланцюжка. Позначимо сигнали на виході лінійного ланцюжка через $y_1(t)$ та $y_2(t)$. Оскільки сигнали $x_1(t)$ та $x_2(t)$ мають зміст логарифмів, то $y_1(t)$ та $y_2(t)$ також будуть логарифмами вихідних сигналів $S_{1вих}(t)$, $S_{2вих}(t)$.

Тоді може бути поставлена обернена задача – перехід від суми $y_1(t)+y_2(t)$ до добутку

$$S_{вих}(t) = S_{1вих}(t)\, S_{2вих}(t). \tag{5.30}$$

Оператор такого перетворення позначимо D^{-1}. Перетворення, обернене логарифмуванню є потенціювання. Характеристика нелінійного елемента, який може здійснювати обернене перетворення, має вигляд

$$\begin{aligned} S_{вих}(t) = D^{-1}(y(t)) = \exp(y_1(t) + y_2(t)) = \exp(y_1(t) \exp(y_2(t) = \\ = \exp(\ln S_{1вих}(t)) \exp(\ln S_{2вих}(t)) = S_{1вих}(t)\, S_{2вих}(t). \end{aligned} \tag{5.31}$$

Схему обробки сигналів в даному прикладі можна подати у вигляді трьох блоків, підключених послідовно. Перший блок виконує нелінійну операцію D, що дозволяє перетворювати добуток сигналів у суму, другий елемент здійснює лінійні операції підсумовування та віднімання. Третій блок, виконуючий операцію D^{-1}, перетворює суму сигналів у добуток.

Якщо сигнали на вході та виході системи розглядати як елементи простору, то перетворення елементів S_1, S_2,…, S_n простору вхідних сигналів в елементи $S_{1вих}$, $S_{2вих}$,…, $S_{nвих}$ простору вихідних сигналів може бути однозначним, але не обов'язково взаємно-однозначним.

Так, наприклад, операція квадрування однозначна, але не взаємно–однозначна.

$$H(S(t)) = (S(t))^2. \tag{5.32}$$

Кожному значенню $S(t)$ відповідає єдине значення $(S(t))^2$, а при оберненому перетворенні отримуємо два можливих значення $\pm S(t)$.

Перетворення векторного простору називається ***гомоморфним***, якщо воно однозначне, але не обов'язково взаємно-однозначне, а системи, що здійснюють таке перетворення, називаються гомоморфними. Якщо перетворення векторного простору однозначне і обернене, то таке перетворення називають ***ізоморфним***.

Логарифмічно-спектральне перетворення привело до окремого напрямку теорії сигналів, який називають ***кепстральним аналізом***. Поняття кепстра визначається як

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln\left(S(\omega) \right)^2 \exp(j\omega q) d\omega, \tag{5.33}$$

де $S(\omega)$ – амплітудний спектр континуального сигналу $S(t)$, який є функцією неперервної змінної t.

Сигнал називають дискретним, якщо він є функцією дискретної змінної, що набуває лише фіксованих значень. $(S(\omega))^2$ представляє спектральну густину енергії сигналу $S(t)$, а $C_s(q)$ можна розглядати як енергетичний спектр функції $\ln(S(\omega))^2$. Аргумент q має розмірність часу, а не частоти. Термін "кепстр" створено переставленням букв у терміні «спектр». q – це особливий кепстральний час, тому що $C_s(q)$ в будь-який момент q залежить від функції $S(t)$, яка задана при $-\infty < t < \infty$. Кепстр, який визначається формулою (5.33), називають ***кепстром потужності***.

Розглянемо отримання кепстра потужності мовного сигналу. Схема обробки сигналу показана на рис. 5.7.

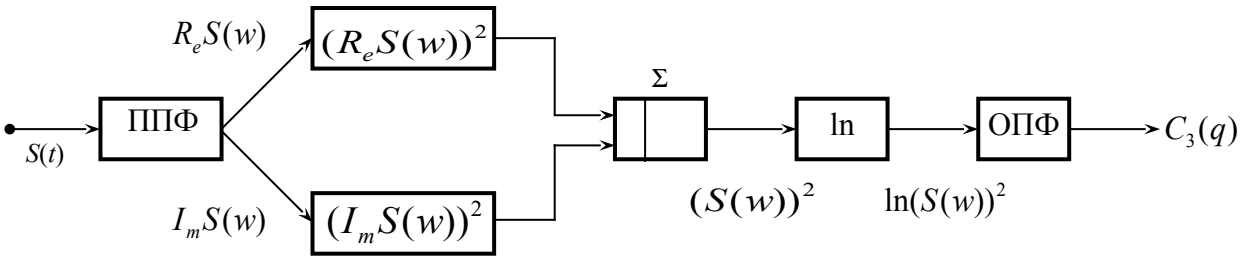


Рисунок 5.7 – Визначення кепстра потужності мовного сигналу:
ППФ – пряме перетворення Фур'є; ОПФ – обернене перетворення Фур'є; Σ – суматор; ln – логарифмуючий пристрій

Мовний сигнал за час $t_2 - t_1$ подають на пристрій, який дозволяє визначити спектр $S(\omega)$ сигналу за допомогою прямого перетворення Фур'є (ППФ).

$$S(\omega) = \int_{t_1}^{t_2} S(t)\exp(-j\omega t)dt = \mathrm{Re}\,S(\omega) - j\Im\, mS(\omega) = S(\omega)\exp(j\theta(\omega)), \qquad (5.34)$$

де

$$\mathrm{Re}\,S(\omega) = \int_{t_1}^{t_2} S(t)\cos\omega t dt, \qquad (5.35)$$

$$\Im\, mS(\omega) = \int_{t_1}^{t_2} S(t)\sin\omega t dt, \qquad (5.36)$$

$$S(\omega) = \sqrt{\left(\mathrm{Re}\,S(\omega)\right)^2 + \left(\Im\, mS(\omega)\right)^2}, \qquad (5.37)$$

$$\theta(\omega) = -\mathrm{arctg}\frac{\Im\, mS(\omega)}{\mathrm{Re}\,S(\omega)}. \qquad (5.38)$$

Вирази (5.35) та (5.36) визначають дійсну й уявну частини спектра. Модуль спектральної густини визначається формулою (5.37), а аргумент – формулою (5.38).

Після зведення в квадрат дійсної та уявної частин спектра проводять їх підсумовування. На виході суматора отримують $(S(\omega))^2$. На виході пристрою, виконуючого обернене перетворення Фур'є, отримують кепстр потужності мовного сигналу $C_s(q)$.

На рис. 5.8 показано кепстр потужності при вимові дзвінкого звука мови.

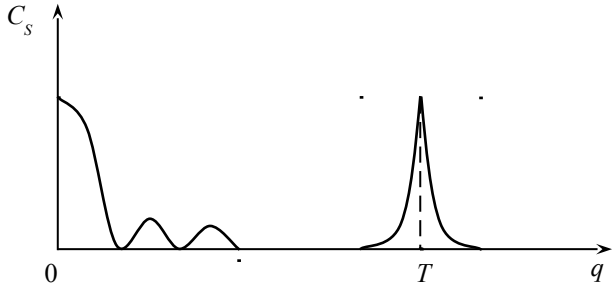


Рисунок 5.8 – Кепстр потужності при вимові дзвінкого звука мови

В області малих значень аргументу q кепстр характеризує мовний тракт, параметри якого повільно змінюються при вимові звуків. Коли $q = T$, де T – період основного тону мови, існує гострий максимум спектра. Максимуми кепстра спостерігаються також там, де значення q відповідає періодам формантних частот.

Схема обробки цифрового сигналу з метою отримання кепстра потужності здійснюється в тій самій послідовності операцій як і при обробці аналогового сигналу. Відліки мовного сигналу в часовій області $S(m)$ піддаються дискретному перетворенню Фур'є (ДПФ) за формулою:

$$S(n) = \sum_{m=0}^{N-1} S(m) \exp(-j\frac{2\pi}{N}mn), \tag{5.39}$$

де $n = 0, 1, \ldots, N-1$; N – число часових відліків сигналу, яке піддається перетворенню Фур'є.

Отримані відліки $S(n)$ являють собою дискретний спектр сигналу. Інтервал між сусідніми спектральними лініями $\Delta\omega$ визначається як

$$\Delta\omega = \pi/NT_д, \tag{5.40}$$

де $T_д$ – період дискретизації сигналу в часовій області. Частота дискретизації $f_д=1/T_д$.

Дискретний спектр також можна подати у вигляді дійсної та уявної складових, тобто

$$S(n) = ReS(n) + j\Im m\, S(n). \tag{5.41}$$

Тому

$$(S(n))^2 = (ReS(n))^2 + (\Im m\, S(n))^2. \tag{5.42}$$

Потім виконують логарифмування і отримують N чисел виду $\ln(S(n))^2$. Застосовуючи обернене дискретне перетворення Фур'є (ОДПФ), визначають кепстр потужності цифрового сигналу

$$C_s(m) = \frac{1}{N}\sum_{n=0}^{N-1} \ln\left(S(n)\right)^2 \exp(j\frac{2\pi}{N}nm), \qquad (5.43)$$

де $m = 0, 1, \ldots, N-1$.

Гомоморфна обробка сигналів здійснюється в гомоморфному вокодері, показаному на рис. 5.9.

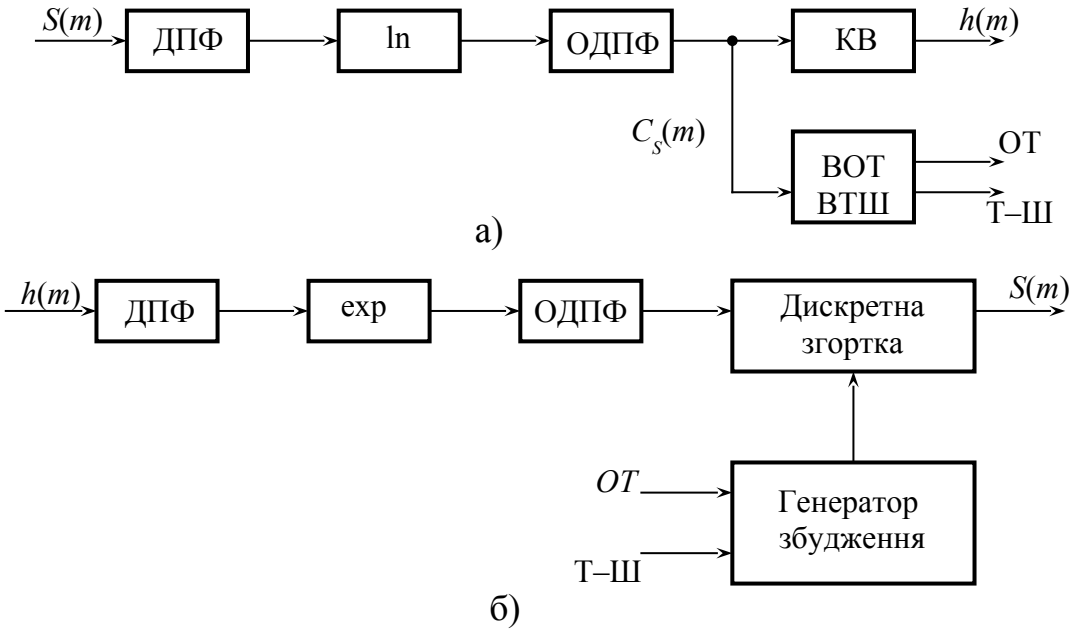


Рисунок 5.9 – Гомоморфний вокодер:
а) аналізатор; б) синтезатор
(ДПФ – дискретне перетворення Фур'є; $S(m)$ – відліки мовного сигналу в часовій області; ln – логарифмуючий пристрій, ОДПФ – обернене дискретне перетворення Фур'є, $C_s(m)$ – кепстр потужності цифрового сигналу, КВ – квантувач, $h(m)$ – компоненти кепстра в області малого часу, ВОТ – виділяч основного тону, ВТШ – виділяч тон-шум, exp – потенціювальний пристрій, ОТ – основний тон, Т-Ш – сигнал тон-шум)

В гомоморфному вокодері кепстр обчислюється через кожні 10…20 мс. За кепстром визначають період основного тону та ознаку тон-шум. Компоненти кепстра в області малих часів (перші 30 відліків кепстра) квантуються за допомогою квантувача і у вигляді параметрів $h(m)$ передаються по каналу зв'язку. В компонентах $h(m)$ міститься інформація про мовний тракт, форму імпульсів збудження та формантну структуру сигналу. Параметри основного тону, сигналу тон-шум та $h(m)$ передаються по каналу зв'язку 50…100 разів за секунду.

У синтезаторі вокодера параметри $h(m)$ піддають дискретному перетворенню Фур'є і отримують логарифм функції, в якій є інформація про мовний тракт, форму імпульсів збудження та формантну структуру. З метою отримання вказаної функції виконують операцію потенціювання за допомогою потенціювального пристрою. Надалі після оберненого дискретного перетворення Фур'є та дискретної згортки відновлюються відліки мовного сигналу в часовій області. Генератором збудження, сигнали якого використовують при дис-

кретній згортці, керують параметри основного тону та сигнали тон-шум, отримані по каналу зв'язку.

Гомоморфний вокодер забезпечує при великому стисненні дуже високу якість звучання мовного сигналу. Недоліком вокодера є складність технічної реалізації. Застосування спеціалізованих процесорів, які здійснюють швидке перетворення Фур'є в реальному масштабі часу, дозволяє суттєво спростити технічну реалізацію гомоморфного вокодера.

5.5.4. Лінійне передбачення мови

Вокодери на основі лінійного передбачення називають ліпредерами. Лінійне передбачення докладно описане в розд. 4.4. Можна вважати, що лінійне передбачення відноситься до найбільш вивчених методів обробки сигналів, тому що воно найпростіше реалізується технічно.

Розроблені ліпредери, в яких використовують ортогональні матричні перетворення коефіцієнтів передбачення. В деяких ліпредерах застосовують кепстральний аналіз при виділенні основного тону мови. Є ліпредери, де проводиться кореляційна обробка сигналів. Також знаходить застосування такий спосіб обробки мови, як зміна частоти передачі відліків. При передачі дзвінких звуків зменшують частоту відліків, а при передачі глухих звуків проводять збільшення частоти відліків. Виявляється, що якість мови помітно покращується, якщо подвоїти частоту відліків при передачі глухих звуків.

Якість мови при використанні лінійного передбачення низька. Не підтвердились припущення про те, що за допомогою ліпредерів можна отримувати мову, за якістю близьку до натуральної. Це пояснюється тим, що при розробці ліпредерів важко врахувати процеси, які відбуваються при мовоутворенні. Робота ліпредерів оцінюється середньоквадратичною помилкою передбачення, при цьому не враховується слухове сприйняття помилок. Модель, що використовується при розробці ліпредерів, відповідає голосним звукам. Ця модель не забезпечує вірний синтез приголосних звуків.

У ліпредерах не можна врахувати взаємне маскування окремих складових спектра мови в критичних смугах слуху, тому що не проводиться спектральний аналіз обвідних звукових сигналів. Передбачення, як метод обробки сигналу може виключити тільки надмірність за рахунок кореляції між відліками сигналу, але не може виключити надмірність, яка спостерігається при слуховому сприйнятті сигналів, тому що при передбаченні не враховуються особливості слуху людини. Тому передбачення не може забезпечити збереження натуральності звучання при суттєвому стисненні мовних сигналів.

Не зважаючи на вказані недоліки, ліпредери широко використовують в сучасній техніці зв'язку. Тому розглянемо їх більш детально.

Згідно з Рекомендацією G.721 Міжнародного союзу Електрозв'язку (МСЕ) при передачі сигналів телефонії з використанням імпульсно-кодової модуляції (РСМ – *Pulse Code Modulation*) швидкість передачі складає 64 кбіт/с. При цьому використовують восьмирозрядне нерів[illegible]квантування і частоту дискретизації 8 кГц. Рекомендацією G.721 МСЕ було передбачено викори-

стання адаптивної диференційної імпульсно-кодової модуляції (ADPCM – *Adaptive Differential Pulse Code Modulation*). Це давало змогу зменшити швидкість цифрового потоку до 32 кбіт/с. Схема кодування мови за Рекомендацією G.721 показана на рис. 5.10 [25].

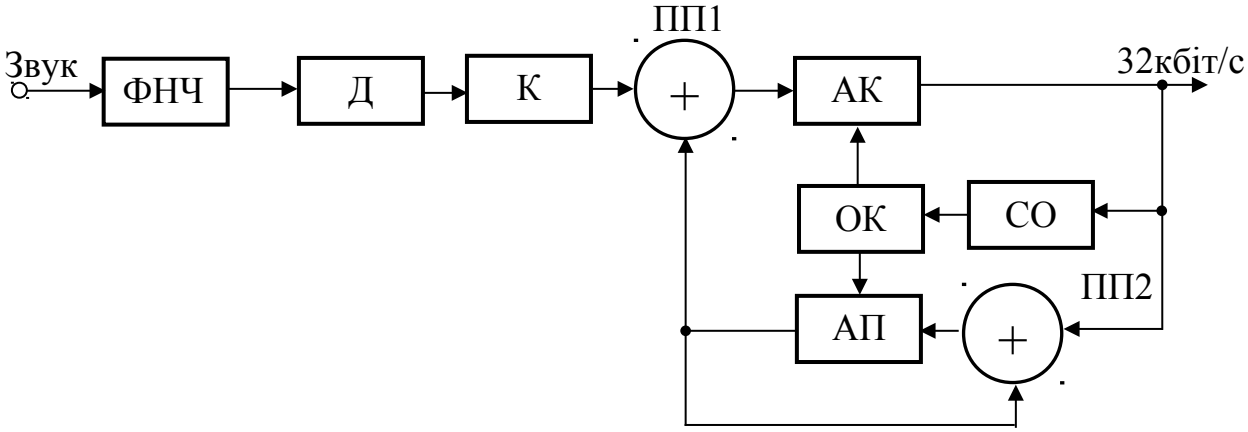


Рисунок 5.10 – Схема кодування мови за Рекомендацією G.721:
ФНЧ – фільтр низької частоти; Д – дискретизатор; К – кодер; АК - адаптивний квантувач; ОК – обчислення коефіцієнтів; СО – статистичне оцінювання; АП – адаптивний передбачник; ПП1…ПП2 – порівнюючі пристрої

На вхід схеми кодування (рис. 5.10) надходять сигнали звуків мови. Фільтр низької частоти (ФНЧ) обмежує сигнал частотою 3,4 кГц. Далі сигнали подаються на дискретизатор (Д), який забезпечує на вході кодера (К) відліки сигналів з частотою 8 кГц. Кодер виконує 12-розрядне лінійне квантування. Потім виконується адаптивне передбачення і адаптивне квантування при чотирьох розрядах на відлік сигналу. В цих операціях беруть участь блоки ПП1, ПП2, на виходах яких отримують помилку порівняння двох сигналів. Блоки ОК та СО забезпечують роботу адаптивного квантувача та адаптивного передбачувача.

Рекомендація G. 721 була прийнята в 1984 році, а через два роки було затверджено стандарт для передачі сигналів зі смугою 50-7000 Гц і швидкістю 64 кбіт/с (Рекомендація G. 722 МСЕ). Схема кодування наведена на рис. 5.11.

У схемі (рис. 5.11) сигнал після дискретизатора і кодера розподіляється цифровими фільтрами на дві смуги (ЦФ1 – смуга 50…4000 Гц, ЦФ2 – смуга 4… 7 кГц). В цій схемі на виході кодера будуть не 12-розрядні, а 14-розрядні відліки при лінійному квантуванні. На вихід адаптивного кантувача АК1 надходять 6-розрядні, а на вихід адаптивного кантувача АК2 – 2-розрядні відліки сигналів. Квантування сигналів на виходах АК1, АК2 – нелінійне [25].

Адаптивні передбачувачі, порівнювальні пристрої, блоки обчислення коефіцієнтів та статистичного оцінювання на рис. 5.11 не показані, але при роботі схеми вони використовуються.

У схемі, рис. 5.11, частота дискретизації складає 16 кГц, але в схемах проріджувачів використовують тільки кожен другий відлік. Тому частота відліків сигналів, які надходять на мультиплексор, дорівнює 8 кГц. Швидкість цифрового потоку на виході мультиплексора – 64 кбіт/с.

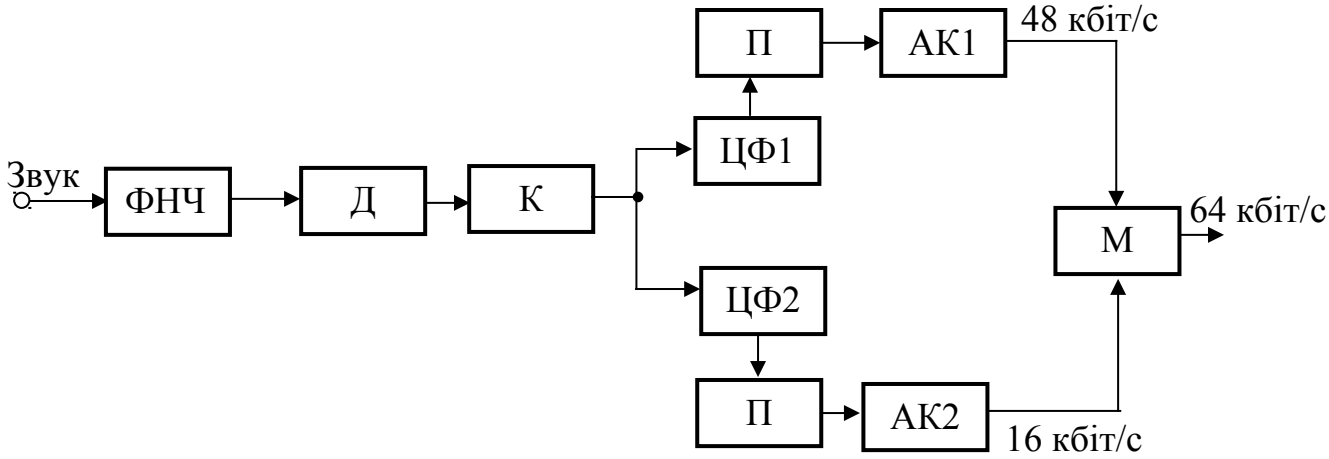


Рисунок 5.11 – Схема кодування мови за Рекомендацією G. 722:
ФНЧ – фільтр низької частоти; Д – дискретизатор; К – кодер; П – проріджувач;
ЦФ1, ЦФ2 – цифрові фільтри; АК1, АК2 – адаптивні квантувачі; М – мультиплексор

ADPCM застосовують також в схемних рішеннях, які відповідають Рекомендації G.726 МСЕ. В них використовують частоту дискретизації 8 кГц та 13-розрядні або 14-розрядні відліки (квантування лінійне). На виході адаптивних квантувачів можуть бути 2-, 3-, 4- або 5-розрядні відліки, отримані при нелінійному квантуванні. Швидкість цифрового потоку, що подається в канал зв'язку може складати 16, 24, 32 або 40 кбіт/с.

У супутникових системах і радіолініях використовують дельта-модуляцію з неперервно змінною крутістю (ДМНЗК), яка поки що не стандартизована МСЕ. Швидкість цифрового потоку складає 12, 16 та 32 кбіт/с. Цей метод відносно недорогий, він реалізований на одній інтегральній схемі.

5.5.4.1. Кодування мови в гібридних кодерах

У гібрідних кодерах використовують комбінацію лінійного передбачення та елементи кодування звукової хвилі. В одному з алгоритмів RELP (Residual Exited Linear Prediction – лінійне передбачення зі збудженням від залишку передбачення) в канал зв'язку передають коефіцієнти лінійного передбачення і підсилення, а також сигнал помилки передбачення в смузі частот 0…800 Гц. Сигнал помилки передбачення передають, використовуючи амплітудне обмеження.

В багатьох гібридних вокодерах застосовують замкнене кодування на основі лінійного передбачення, яке ще називають ***методом аналізу через синтез***. Таке кодування дозволяє знаходити найкращу апроксімацію сегмента мовного сигналу. Як тільки апроксімація буде знайдена, її код передають на приймальну сторону каналу зв'язку. Цей код використовується при синтезі сигналів.

Метод «аналізу через синтез» використовується в алгоритмі лінійного передбачення з багатоімпульсним збудженням MPE-LPC (Multi Pulse Excitation

LPC). Тут і далі LPC (Linear Predictive Coding) – кодування на основі лінійного передбачення [25].

При роботі алгоритму MPE-LPC виконується мінімізація помилки передбачення $e(n)$ (рис. 5.12).

На передавальній стороні проводять синтез сигналу $\overline{S}(n)$ на основі вхідного сигналу $S(n)$ і шляхом мінімізації різниці між ними. При цьому підбирають відповідні сигнали збудження. Мінімізують також енергію зваженої помилки передбачення $e_w(n)$.

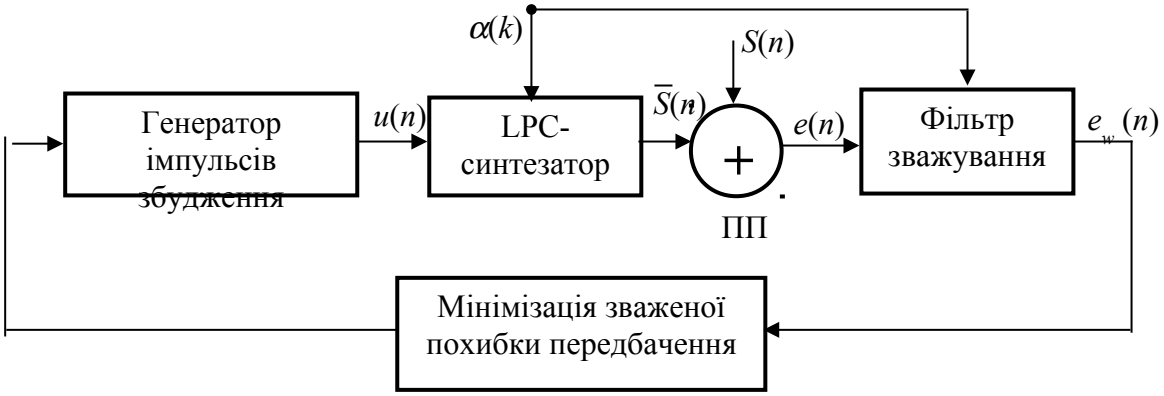


Рисунок 5.12 – Схема визначення сигналу багатоімпульсного збудження:
ПП – порівнювальний пристрій; $u(n)$ – сигнал збудження; $\alpha(k)$ – коефіцієнти лінійного передбачення; $S(n)$ – синтезований сигнал; $S(n)$ – вхідний сигнал; $e(n)$ – помилка передбачення; $e_w(n)$ – зважена похибка передбачення

Кодери MPE-LPC забезпечують швидкість передачі 9,6 і 16 кбіт/с, причому якість синтезованої мови відповідає якості мови, яку отримують у звичайних ІКМ-системах при швидкості 52 і 58 кбіт/с.

Якість передачі підвищується, коли використовують довгочасні надмірності. Ці кодери називають LTP (Long Term Predictor) – довчасні передбачувачі.

В одній різновідності алгоритмів приміняють регулярне імпульсне збудження RPE (Regular Pulse Excitation), за якого сигнал збудження являє собою послідовність рівномірно розміщених імпульсів. Алгоритм (RPE-LTP-LPC) регулярного імпульсного збудження з довгочасним передбаченням використовують у сотових системах зв'язку при швидкості цифрового потоку 13 кбіт/с. Якість телефонного зв'язку за такої швидкості наближається до якості телефоних розмов у мережах загального користування.

Алгоритм лінійного передбачення з довгочасним передбаченням та багатоімпульсним збудженням (MPE-LTP-LPC) застосовують у сучасних супутникових системах зв'язку.

У 1985 році був запропонований алгоритм лінійного передбачення з кодовим збудженням CELP (Code Excited Linear Prediction), який забезпечує якісну передачу за малої швидкості цифрового потоку [25].

Синтезатор декодера CELP являє собою фільтр, який формує до чотирьох формант у смузі частот стандартного телефонного каналу. Періодична структура сигналу збудження моделюється довчасним ситензувальним

фільтром, час затримки якого складає від 20 до 120 відліків сигналу. Це відповідає частоті основного тону мови від 75 до 400 Гц. Короткочасний фільтр (КФ) дозволяє формувати до чотирьох формант в смузі каналу зв'язку.

Кодери і декодери CELP мають кодову книжку (КК), з якої вибирають варіанти сигналів. З КК аналізатор вибирає оптимальний сигнал, який забезпечує найкраще приближення синтезованого сигналу $\overline{S}(n)$ до вхідного сигналу $S(n)$ в відповідності з критерієм мінімуму енергії зваженої помилки передбачення $E(i)$.

Робота алгоритму CELP зводиться до мінімізації енергії зваженої помилки передбачення. Фільтр зважування розподіляє шуми квантування за частотним діапазоном таким чином, щоб енергія цих імпульсів мало впливала на якість мови. Це буде в тому випадку, коли шуми квантування будуть знаходитись в формантних областях.

Пост-фільтр (ПФ) перерозподілює енергію шуму між формантами при синтезі, що покращує якість звучання мови.

Схема кодера і декодера CELP наведена на рис. 5.13.

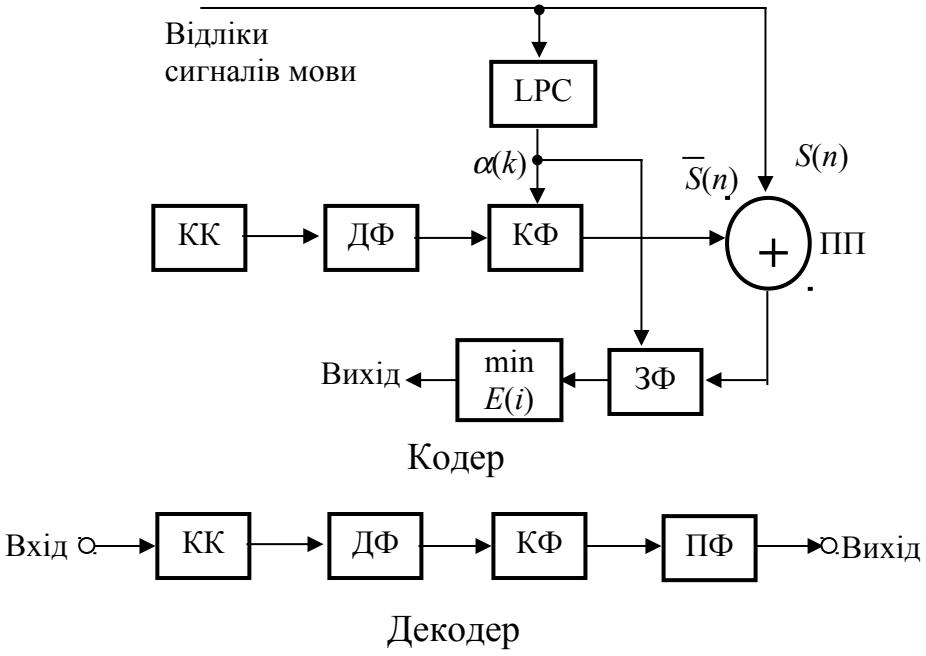
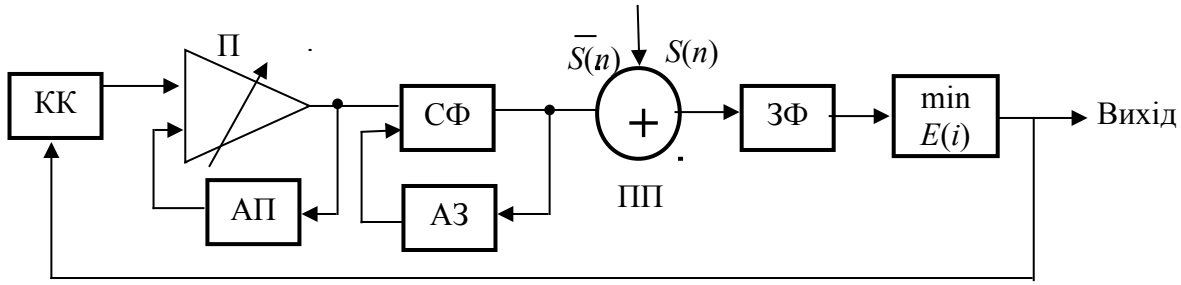


Рисунок 5.13 – Схема кодера та декодера CELP:
ДФ – довгочасний синтезувальний фільтр; КФ – короткочасний синтезувальний фільтр; КК – кодова книга; ПП – порівнюючий пристрій; LPC –аналізатор на основі лінійного передбачення; ЗФ – зважувальний фільтр; $E(i)$ – мінімізатор енергії зваженої помилки; $\alpha(k)$ – коефіцієнт лінійного передбачення; $\overline{S}(n)$ – синтезований сигнал; $S(n)$ – вхідний сигнал; ПФ – пост-фільтр

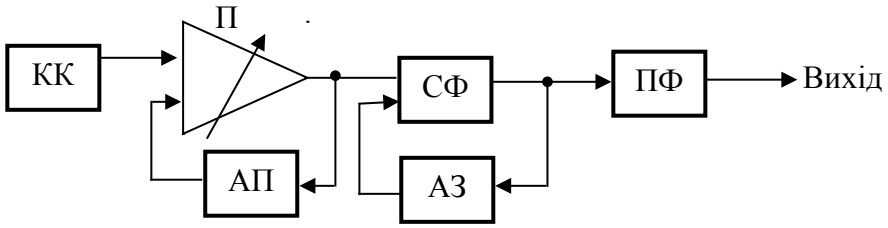
ЗФ відіграє ту саму роль, що і в алгоритмі MPE-LPC.

Алгоритм лінійного передбачення з кодовим збудженням і малою затримкою LD-CELP (Low delay CELP) забезпечує кодування 16-розрядним лінійним кодом при швидкості передачі 16 кбіт/с і частоті дискретизації 8 кГц. Використовуються цифрові сигнальні процессори, затримка алгоритму складає не більше 2 мс [25].

На рис. 5.14 показана схема кодера та декодера LD-CELP.



Кодер



Декодер

Рисунок 5.14 – Схема кодера та декодера LD-CELP:
КК – кодова книга; П – підсилювач; АП – адаптація підсилення; СФ –синтезувальний фільтр; АЗ – адаптація передбачення; $\bar{S}(n)$ – синтезований сигнал; $S(n)$ – вхідний сигнал; ПП – порівнювальний пристрій; ФЗ – фільтр зважування; min $E(i)$ – мінімізатор зваженої середньоквадратичної похибки; ПФ – пост-фільтр

Синтезувальний фільтр має тільки короткочасний передбачувач для підвищення якості звучання високих голосів мови. Коефіцієнт передбачення не передають з метою досягнення малого часу затримки.

З кодової книги вибирають сигнал збудження для кожних з п’яти відліків сигналу. Параметр сигналу збудження містить номер сигналу збудження в кодовій книзі, знак сигналу та коефіцієнт підсилення. В кодовій книзі розміщується 256 сигналів збудження. Коефіцієнт підсилення призначений для регулювання інтенсивності синтезованої мови.

У кодері LD-CELP виконується оптимізація сигналу збудження, який беруть з кодової книги, з використанням фільтра зважування та мінімізатора зваженої середньоквадратичної похибки. Оптимізація підвищує якість звучання синтезованої мови.

Пост-фільтр використовується з тією самою метою, що і в алгоритмі CELP.

У листопаді 1991 року почались роботи над двома алгоритмами CELP: ACELP (Alqebraic CELP – лінійне передбачення зі збудженням алгебраїчним кодом) та CS-CELP (Conjugate Structure CELP – лінійне передбачення з кодовим збудженням з’єднаної структури) [25].

Алгоритм ACELP використовується у Франції та Канаді, а алгоритм CS-CELP – в Японії.

В алгоритмі ACELP коефіцієнт передбачення визначають 100 раз в секунду, інтервал аналізу розбивається на два підсегменти. Параметри довгочасного передбачувача знаходять за допомогою адаптивної кодової книжки. Використовують також критерії мінімізації довгочасної похибки. Кодова книга містить алгебраїчні коди, які являють собою блок з 40 відліків.

Алгоритм ACELP і CS-CELP забезпечують швидкість передачі 8 кбіт/с, використовуються в сотових системах зв'язку та супутникових системах передачі. Існує версія ACELP з швидкістю передачі 5,4 кбіт/с.

У системах зв'язку Франції використовують також алгоритм CS-ACELP, який забезпечує швидкість передачі 8 кбіт/с і відповідає Рекомендації G.729 ITU-T.

5.5.4.2. Малошвидкісні алгоритми та кодування мови з змінною швидкістю

В основі малошвидкісних алгоритмів лежать принципи лінійного передбачення. В цих алгоритмах використовують різні методи передачі сигналів збудження, коефіцієнтів передбачення та способів кодування передаваних параметрів. Сигнали збудження передають з періодом основного тону мови при вокалізованих сегментах. У випадку невокалізованих сегментів сигнали збудження являють собою випадкову послідовність імпульсів.

Цей метод лежить в основі алгоритму LPC-10 (Linear Predictive Coding-10 – кодування на основі лінійного передбачення), який прийнято в США для кодування мови з швидкістю 2,4 кбіт/с. Стандарт, який визначає алгоритм LPC-10, містить вимоги до кодера та декодера мови.

Більш складний алгоритм MPE-LTP-LPC (Multi Pulse Excitation-Long Term Predictor-LPC – лінійне передбачення з довчасним передбаченням і багатоімпульсним збудженням) вимагає більшої швидкості цифрового потоку і забезпечує високу якість синтезованої мови. Алгоритм рекомендовано як міжнародний стандарт для мобільного, авіаційного та супутникового зв'язку. Швидкість передачі в алгоритмі MPE-LTP-LPC складає 9,6 кбіт/с [25].

Тип сегмента – вокалізований (тональний), невокалізований (шумовий), паузу в алгоритмі LPC-10 визначають на основі аналізу енергії нижньої смуги частот мовного сигналу, за значеннями кореляції при затримці, що дорівнює періоду основного тону, на частоті переходів сигналу через нуль та за крутістю сигналу. Використовується також завадостійке кодування з метою забезпечення надійності алгоритму за високих значень ймовірності помилок при передачі сигналів.

Вхідний сигнал $S(n)$ подається на смуговий фільтр з смугою 100…3600Гц. Далі сигнал надходить на аналого-цифровий перетворювач з частотою дискретизації 8000 Гц і 12-розрядним лінійним кодуванням. Потім виконується передспотворення оцифрованого сигналу блоком БП. Передспотворення дозволяють знизити вимоги до інтегральної точності оцінок параметрів передбачення.

Структурна схема кодера LPC-10 наведена на рис. 5.15.

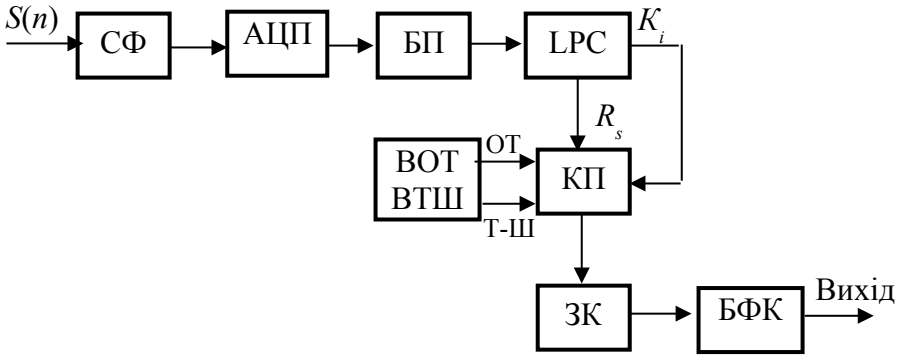


Рисунок 5.15 – Схема кодера LPC-10:
$S(n)$ – вхідний сигнал; СФ – смуговий фільтр; АЦП – аналого-цифровий перетворювач; LPC – LPC-аналізатор; K_i – коефіцієнт відбиття; R_s – характеристика енергії сигналу на інтервалі аналізу; ВОТ, ВТШ – виділячі основного тону і сигналу тон-шум; КП – кодування параметрів; ЗК – завадостійке кодування; БФК – блок формування кадрів; БП – блок передспотворення; ОТ – сигнал основного тону; Т-Ш – сигнал тон-шум.

За допомогою LPC-аналізатора знаходять коефіцієнти відбиття K_i та характеристику енергії сигналу на інтервалі аналізу R_s . Коефіцієнти відбиття та коефіцієнти передбачення зв'язані між собою матрицею. Характеристика енергії сигналу на інтервалі аналізу являє собою середньоквадратичну амплітуду відліків сигналів на інтервалі аналізу.

Інтервал аналізу складається з 130 відліків, довжина сегмента аналізу 22,5 мс. На передачу кожного сегмента виділяється 54 біт, швидкість передачі 2400 біт/с. 54 біт розглядається як кадр. 7 біт виділяється для передачі сигналу основного тону і сигналу тон-шум, 5 біт – для передачі характеристики енергії сигналу R_s. Для передачі коефіцієнтів відбиття виділяють від 2 до 5 біт [25].

Тип сегмента (вокалізований чи невокалізований) передають двічі за кадр. При цьому можливі такі чотири типи сегментів:

1) 00 – невокалізований (шумовий) стан;

2) 01 – перехід від невокалізованого до вокалізованого (тонального) стану;

3) 10 – перехід від вокалізованого до невокалізованого стану;

4) 11 – вокалізований стан.

У випадку вокалізованого стану визначають період основного тону мови. Якщо сегмент невокалізований, то визначають чи це шумовий сегмент, чи в даний інтервал існує пауза мови.

Коли передається невокалізований сегмент, то 20 біт виділяється для завадостійкого кодування. Якщо передають вокалізований сегмент, то завадостійке кодування не виконують. Один біт в кадрі призначено для синхронізації.

Блок формування кадрів формує 54-бітові кадри. Схема декодера LPC-10 показана на рис. 5.16.

Блок корекції помилок находить помилки і виправляє їх. Декодовані декодером параметри надходять в блок інтерполяції. Параметри інтерполюються на основі їх значень в попередньому і наступному кадрах, тобто відбувається передбачення «вперед» і «назад». Для цього реалізована затримка сигналів

на один кадр. Зі збільшенням кількості помилок збільшується ступінь згладжування параметрів.

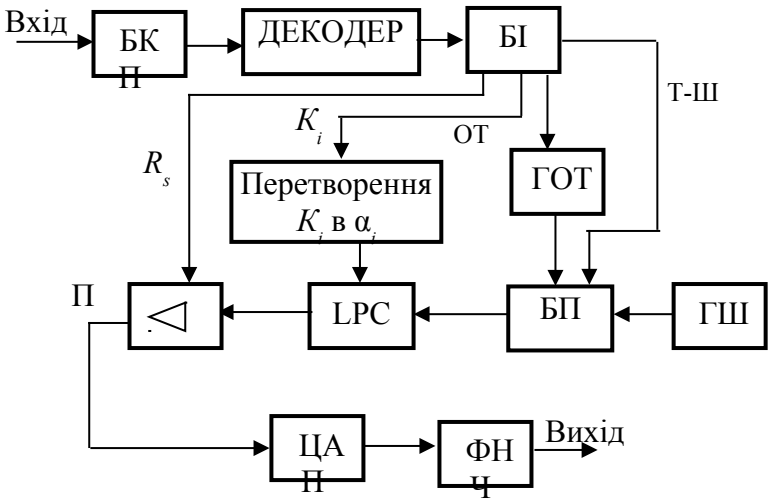


Рисунок 5.16 – Схема декодера LPC-10:
БКП – блок корекції помилок; БІ – блок інтерполяції; R_s – характеристика енергії сигналу на інтервалі аналізу; K_i – коефіцієнт відбиття; ОТ – сигнал основного тону; Т-Ш – сигнал тон-шум; ГОТ – генератор основного тону; П – підсилювач; LPC – LPC-аналізатор; БП блок переключення; ГШ – генератор шуму; ЦАП –цифро-аналоговий перетворювач; ФНЧ – фільтр низької частоти

Коефіцієнти K_i інтерполюються двічі за кадр. Параметр R_s інтерполюється для кожного періоду основного тону мови. Протягом всього сегмента інтерполюється період основного тону мови. Якщо один за одним ідуть вокалізований та невокалізований сегменти, то в цьому випадку коефіцієнти K_i не інтерполюються.

Коли в даний момент часу передають вокалізований сегмент, то на виході блока переключення будуть сигнали генератора основного тону. Якщо передають невокалізований сегмент, то на вихід блока переключення надходять сигнали генератора шуму.

Інтерпольовані значення коефіцієнтів відбиття K_i перетворюються в значення коефіцієнтів передбачення α_i, які далі використовуються в LPC-синтезаторі. Сигнали збудження на виході LPC-синтезатора мають постійну величину, тому вони масштабуються в підсилювачі П за допомогою сигналів R_s.

Після цифро-аналогового перетворювача сигнали подаються на фільтр низької частоти.

Для вокалізованих сегментів вибирають одне із 60 значень частоти основного тону мови з діапазону частот 51…400 Гц.

Завадостійке кодування виконують тільки для невокалізованих сегментів і застосовують його для захисту чотирьох старших розрядів параметра R_s і перших чотирьох коефіцієнтів відбиття.

В алгоритмі LPC-LTP-MPE використовують частоту дискретизації 8000 Гц і 12-розрядний лінійний код. Послідовність вхідних відліків сигналів розбивається на такі сегменти: розширений – 256 відліків; основний – 160 відліків.

Кожний основний сегмент розбивається на п’ять підсегментів по 32 відліки в кожному піделементі [25].

У кодері LPC-LTP-MPE (див. рис. 5.17) виділяють такі параметри: 10 коефіцієнтів лінійного передбачення (α_1,…, α_{10}), які зв’язані з коефіцієнтами відбиття ($K_1, ..., K_{10}$); коефіцієнти довгочасного передбачення – затримка і підсилення; 30 параметрів сигналів багатоімпульсного збудження (15 положень імпульсів збудження ($m_1, ..., m_{15}$) і 15 значень їх амплітуд ($g_1, ..., g_{15}$).

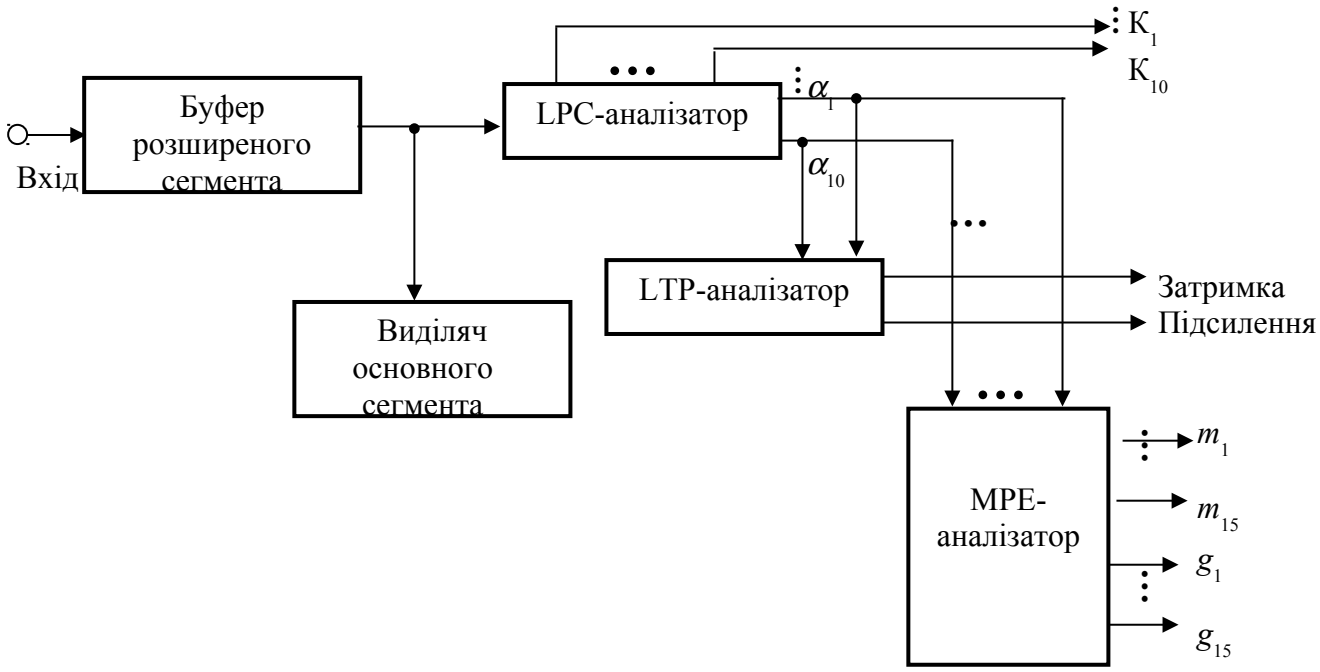


Рисунок 5.17 – Кодер MPE-LTP-LPC:

$K_1,..., K_{10}$ – коефіцієнти відбиття $\alpha_1, ..., \alpha_{10}$ – коефіцієнти лінійного передбачення; $m_1,..., m_{15}$ – положення імпульсів багатоімпульсного збудження; $g_1, ..., g_{15}$ – амплітуди імпульсів багатоімпульсного збудження

Параметри «Затримка» і «Підсилення» виділяють на основному сегменті, а параметри MPE-аналізатора – на кожному підсегменті. При виділянні параметрів використовують автокореляційні функції. Сигнали багатоімпульсного збудження є послідовність з трьох імпульсів з нерівномірно розподіленими інтервалами і різними амплітудами.

Для передачі основного сегмента виділяють 192 біти. Завадостійкість передачі забезпечується кодом Хеммінга, на який припадає 14 біт. Для кодування коефіцієнтів відбиття виділяють від 2 до 6 біт, для кодування параметра «Затримка» – 6 біт, для кодування параметра «Підсилення» – 2 біти, для кодування імпульсів багатоімпульсного збудження – від 3 до 5 біт. Значення кодованих параметрів об’єднують у кадри величиною 192 біти.

Декодер LPC-LTP-MPE являє собою три синтезатори, які включені послідовно. Спочатку синтезуються імпульси збудження (MPE-синтезатор). Потім сигнали надходять на LTP-синтезатор і далі – на LPC-синтезатор.

Розглянуті алгоритми забезпечують передачу мовних сигналів з незмінною швидкістю. Нові телекомунікаційні технології успішно справляються з нерівномірним трафіком. Тому були розроблені алгоритми кодування мови зі змінною швидкістю цифрового потоку.

Методи лінійного передбачення широко використовують в мережах зв'язку, а також для зберігання, обробки і передачі сигналів мови. Але при передачі мови зі швидкістю 4 кбіт/с і менше, якість мови різко погіршується.

Можливість передачі мови з змінною швидкістю забезпечується використанням технологій комутації пакетів, протоколів з змінними швидкостями, високошвидкісних мультиплексорів потоків зі змінними швидкостями. Існує думка, що змінна швидкість передачі буде головним напрямом розвитку цифрових мереж у майбутньому.

Кодування мови зі змінною швидкістю ґрунтується на класифікації сегментів мовного сигналу і використанні різних систем кодування на різних сегментах мови.

Передача сигналів мови з постійною швидкістю може виконуватись в мережах з комутацією каналів і в мережах з комутацією пакетів. Для передачі сигналів мови з постійною швидкістю можуть використовуватись алгоритми передачі сигналів мови зі змінною швидкістю.

Висока ефективність мереж зв'язку передачі зі змінною швидкістю буде тоді, коли використовують колективний доступ з кодовим розподілом (Code Division Multiple Access, CDMA). Складовою частиною CDMA є модуль контролю змінної швидкості (Variable Rate Control Unit, VRCU), який забезпечує оптимальний розподіл смуги пропускання каналу зв'язку між різними джерелами сигналів [25].

Модуль VRCU аналізує сигнали, які подаються на мультиплексор, і виконує розподіл смуг пропускання. При цьому модуль аналізує:

– потреби джерел інформації;

– потреби і можливості мережі зв'язку;

– вимоги споживачів інформації.

Структура кодера змінної швидкості на основі фонетичної класифікації наведена на рис. 5.18.

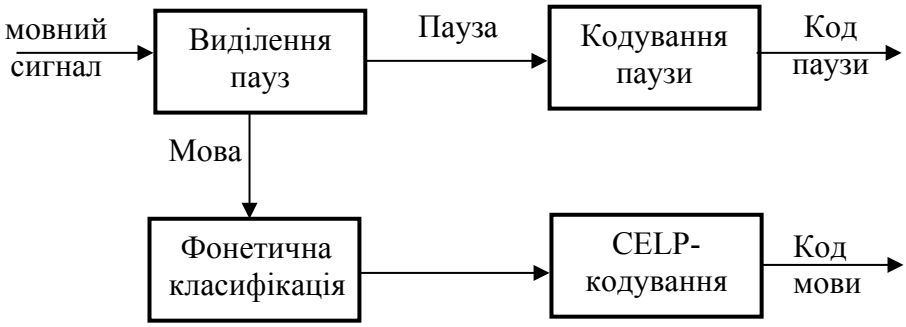


Рисунок 5.18 – Узагальнена схема кодування зі змінною швидкістю на основі фонетичної класифікації

При класифікації виділяють декілька фонетичних типів, які відповідають різним величинам ентропії сигналу. В залежності від класифікації вибирають схему кодування даного сегмента мови.

Основою класифікації є ознака «вокалізований/невокалізований». Сегменти мови довжиною 20 мс ділять на чотири підсегменти довжиною по 5 мс і визначають тип підсегмента. Визначення вокалізованості підсегмента відбувається на основі аналізу таких параметрів:

– нормалізованої енергії в смузі нижніх частот;

– нормалізованої енергії в смузі верхніх частот;

– частоти переходу сигналу через нуль;

– величини першого і другого коефіцієнтів відбиття.

Класифікація сегмента відбувається на основі аналізу підсегментів. При цьому визначають чи вокалізований, чи невокалізований підсегмент. Для кожного фонетичного класу існує відповідний алгоритм при виконанні наступних дій: 1) обчисленні і кодуванні параметрів сигналів збудження; 2) виконанні LPC-аналізу і квантуванні коефіцієнтів лінійного передбачення; 3) обчисленні зваженої помилки передбачення; 4) адаптивній постфільтрації сигналів збудження у випадку передачі вокалізованих сегментів.

Блок, який вибирає сигнали збудження, по різному працює при передачі вокалізованих, невокалізованих чи невизначених підсегментів. Використовують в залежності від типу підсегмента, різні розміри кодових книг і по різному кодуються коефіцієнти підсилення. У випадку вокалізованих сегментів передають параметри довгочасного передбачення сигналів збудження.

Швидкість цифрового потоку при передачі паузи складає 0,75 кбіт/с, при передачі невокалізованого сегмента – 2,35 кбіт/с, при передачі вокалізованого сегмента – 5,75 кбіт/с. Середня швидкість передачі – нижче 3 кбіт/с.

Мова при використанні алгоритму зі змінною швидкістю була якісна і вільна від реверберацийних спотворень, які характерні для низькошвидкісних кодерів CELP. Якість мови була не гіршою ніж при використанні алгоритму CELP з постійною швидкістю цифрового потоку величиною 4,8 кбіт/с.

Розглянемо алгоритм кодування зі змінною швидкістю, який реалізовано на однокристальній мікросхемі Q 4401. Цей алгоритм кодує мовні сигнали в режимі постійної швидкості (4; 4,8; 8; 9,6 кбіт/с) і змінної швидкості в діапазоні 800-9600 біт/с [25].

Швидкість передачі регулюється кожні 20 мс. Середня швидкість передачі при неперервній мові складає 7 кбіт/с, а при звичайному двосторонньому телефонному зв’язку – 3,5 кбіт/с.

Програма алгоритму зберігається в запам’ятовувальному пристрої цифрового сигнального процесора. Управління роботою мікросхеми Q 4401 виконує цифровий сигнальний процесор. Алгоритм кодування змінюється в залежності від енергії сигналу в сегменті. Якщо енергія сигналу максимальна, то використовується максимальна швидкість передачі. При середній енергії си-

гналу буде проміжне значення швидкості цифрового потоку, а при малій енергії сигналу швидкість цифрового потоку складає 800 біт/с.

При нормальному режимі змінної швидкості та максимальної енергії сигналу швидкість цифрового потоку складає 8000 біт/с. Якщо використовують розширений режим змінної швидкості, то максимальній енергії сигналу буде відповідати швидкість 9600 біт/с. При звичайній телефонній розмові середня швидкість складає 6 кбіт/с, а якість розмови відповідає стандарту G.728.

Запропоновано алгоритм змінної швидкості, який забезпечує постійну якість синтезованої мови при середній швидкості 6 кбіт/с і максимальній швидкості 16 кбіт/с. В цьому алгоритмі аналізуються сегменти у 80 відліків сигналу. Кожен сегмент ділиться на підсегменти у 20 відліків сигналу. В аналізаторі виділяють параметри, які відносять як до сегмента, так і до підсегментів. Вибором алгоритму обробки сигналів керують блок класифікації по мовному сигналу і блок класифікації методом аналізу через синтез.

Основним елементом алгоритму, який забезпечує постійну якість мови, є багатошвидкісний кодер CELP. Блок класифікації по мовному сигналу виділяє параметри, які найкращим чином представляють текучі мовні сегменти, і підмножину можливих швидкостей передачі. Блок класифікації методом аналізу через синтез дозволяє знаходити найменшу швидкість, яка відповідає необхідній якості синтезованої мови.

В багатошвидкісному кодері CELP використовується вісім швидкостей: 0; 0,4; 3,2; 8,5; 12,5; 7,2; 12 та 16 кбіт/с. Перші три швидкості призначені для кодування фонового шуму, а решта – для кодування мови. Останні п’ять режимів забезпечуються шляхом додаткової передачі параметрів збудження. Коефіцієнти короткочасного передбачення знаходять за допомогою автокореляційного методу. Потім ці коефіцієнти перетворюють в лінійні спектральні пари.

Параметри довгочасного передбачення являють собою затримку і коефіцієнт довгочасного підсилення. Їх знаходять в інтервалі між 20 та 147 відліках сигналів. Для оцінки періоду основного тону мови використовують автокореляційну функцію.

Числові значення затримки і підсилення знаходять шляхом мінімізації енергії зваженої похибки передбачення між мовним сигналом і результатом довгочасного передбачення мовного сигналу. Для визначення зваженої похибки використовують адаптивний інтерполятор, який забезпечує високу точність за малого числа обчислень.

Багатошвидкісний кодер CELP аналізує три групи параметрів збудження:

1) параметри довгочасного передбачення;

2) параметрів сигналу збудження А (кодова книга А сигналів збудження і підсилення);

3) параметрів сигналу збудження В (кодова книга В сигналів збудження і підсилення). Виконується оптимізація параметрів методом аналізу через синтез. При цьому параметри послідовно включають в процедуру аналізу через синтез до тих пір поки величина похибки синтезу не стане меншою за певний поріг.

Квантування параметрів підсилення виконується логарифмічним квантувальним пристроєм з кроком 3 дБ.

Для кожного сегмента мови виконується фонетична класифікація: визначають чи це пауза, чи це мова; вокалізований сегмент, чи невокалізований. Якщо з'ясували, що сегмент в стані «мова», то потім класифікують вокалізований сегмент чи невокалізований. Сегмент вважають вокалізованим тоді, коли затримка довгочасного передбачення та логарифмічний коефіцієнт підсилення перевищує певні пороги або дорівнюють їм.

Значення порогів залежать від фонового шуму і мови абонента. Адаптивні пороги знаходять, враховуючи параметри довгочасного передбачення. Числові значення адаптивних порогів обмежені певними величинами.

Середня швидкість цифрового потоку при використанні багатошвидкісного кодера CELP змінюється в межах від 6,1 до 7,6 кбіт/с. Якість синтезованої мови близька до стандарту G. 728.

У табл. 5.3 наведена порівняльна характеристика алгоритмів кодування мови, які використовують лінійне передбачення. Там же приведено стандарти і використання алгоритмів [25].

Таблиця 5.3 – Порівняльна характеристика алгоритмів кодування мови

Алгоритм кодування мови	Швидкість, кбіт/с	Стандарт	Використання
PCM	64	ITU-T G.711	Громадські телефонні мережі
ADPCM	32	ITU-T G.721	Громадські телефонні мережі, зберігання мови, цифрові безпроводні телефони
LD-CELP	16	ITU-T G.728	Громадські телефонні мережі, відеотелефон
RPE-LTP	13	ETSI GSM	Європейська цифрова сотова мережа
MPE-LTP	9.6	AEE C	Супутникова система зв'язку INMARSAT AERONAUTIKAL
CS-ACELP	8	ITU-T G.729	Передача мови в мережах FRAME RELAY, ATM, в системах телезв'язку Франції
VSELP	8; 5.6	TIA IS54	Цифрові сотові системи США
CELP	4.8	ANSI	Спеціальні системи зв'язку
MP-MLQ	4.8; 6.4; 7.2; 8.0; 5.3; 6.3	ITU-T G.723	Системи зв'язку. Рекомендація H.324 – передача мови в відеотелефонах
LPC	2.4	ANSI	Спеціальні системи зв'язку

У складних алгоритмах обробки мовних сигналів збільшуються часові затримки, які знижують якість телефонних розмов. У відповідності з Рекомендацією ITU-T G.114 затримка мовного сигналу на 150 мс вважається прийнятною, а затримка в 400 мс – недопустимою. В розглянутих алгоритмах часові затримки такі: PCM, ADPCM – 125 мкс; LD-CELP – 5 мс; MPE-LTP – 34

мс; CS-ACELP –38 мс; VSELP – 68 мс; CELP – 105 мс; MP-MLQ – 22 мс; LPC-10 –135 мс.

З наведених даних видно, що в усіх розглянутих алгоритмах лінійного передбачення мовних сигналів часові затримки прийнятні і відповідають Рекомендації G.114 ITU-T.

На рис. 5.19 показана діаграма якості звучання мови за п’ятибальною шкалою при використанні різних алгоритмів кодування.

Внизу рис. 5.19 наведені швидкості цифрового потоку в кбіт/с, які відповідають алгоритмам кодування мови. Найвищу оцінку (4,5 бала) має алгоритм ADPCM, при швидкості цифрового потоку 64 кбіт/с (Рекомендація G.722.1 ITU-T). Найнижча оцінка в алгоритмі LPC-10.

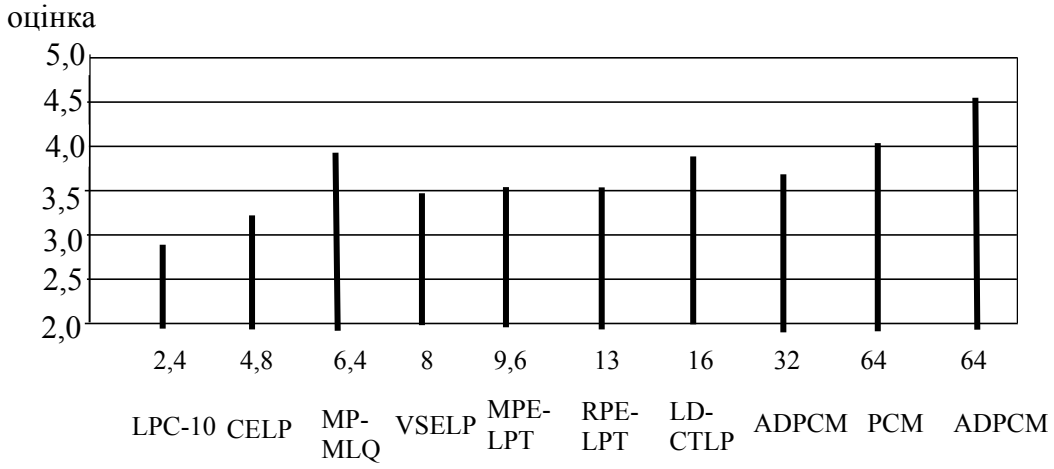


Рисунок 5.19 – Діаграма якості звучання мови за п’ятибальною шкалою

Наведені дані свідчать, що ліпредери мають значно гірші показники якості передачі ніж ті, що забезпечуються смуговими та формантними вокодерами.

5.6. Мовоелементні методи

З усіх розглянутих методів більш близькі до ідеального стиснення мовного сигналу мовоелементні методи. При цьому необхідна пропускна здатність каналу зв’язку може бути майже такою, як і при телеграфній передачі.

У середньому швидкість вимови складає близько 10 фонем за секунду. Мова містить інформацію про індивідуальність абонента, відображає його емоції та настрій. Якщо залишити в мові тільки інформацію про вимовляні фо-

неми, а решту інформації, яка характеризує голос абонента, вилучити, то необхідна пропускна здатність каналу зв'язку для такої передачі буде дуже малою. Російська мова містить 41 фонему, тому при шестизначному коді ($2^6 = 64$) можна передати не лише фонеми, але й іншу інформацію про голос. Необхідна пропускна здатність каналу зв'язку буде $10 \times 60 = 60$ біт/с. Отже, величина стиснення фонемного вокодера може бути в межах 1000. При цьому втрачається індивідуальність голосу і таку передачу називають звуковим телеграфом.

Існують два напрямки розробки мовоелементних вокодерів. Перший напрямок передбачає розпізнавання елементів мови в аналізаторі, кодування та передачу їх по каналу зв'язку. Відновлення здійснюється за зразками елементів мови, що зберігаються в пам'яті синтезатора. Якість мови при відновленні низька, оскільки є стрибки в місцях стику елементів мови.

При реалізації методів другого напрямку визначають частоту основного тону мови, рівень інтенсивності та тривалість звучання кожного елемента мови. Ці дані кодуються і передаються по каналу зв'язку. Синтезатор використовує передані по каналу зв'язку параметри для відновлення мови. Також інколи передають дані про частоти та смуги формант. Якість мови при цьому суттєво покращується [6].

В мовоелементних вокодерах при передачі стискують мовний сигнал з метою зменшення пропускної здатності каналу зв'язку. В синтезаторі фонемного вокодера не тільки проводять експандування сигналу, але й керують процесом перетворення кодових сигналів у мовний сигнал. Вокодери, розглянуті раніше, в основному, тільки виконували компресію сигналу при передачі та експандування при прийомі.

Мовоелементні методи аналізу та синтезу сигналів можуть використовуватись не тільки для організації вокодерного телефонного зв'язку, але й використовуватись при вирішенні інших задач. Так, наприклад, при керуванні машинами та технологічними процесами за допомогою голосу нема необхідності копіювати слуховий аналізатор людини, оскільки голос сприймає машина. Важливо забезпечити правильне розпізнавання звуків мови, окремих слів та речень. Якщо мова фіксується автоматичною друкарською машиною, то треба за вимовленими словами вірно визначити їх друковане подання. Якщо необхідно вірно відтворити речення, що складені самою машиною або при сприйманні машиною надрукованого тексту, то треба враховувати слухові властивості вуха людини та правила вимови.

Мовоелементні методи перетворення мовних сигналів дуже різноманітні і використовуються в багатьох галузях. Нижче будуть розглянуті тільки методи, що застосовуються у техніці вокодерного зв'язку.

При фонемному кодуванні враховують три етапи сприйняття мови слуховим аналізатором людини. В процесі першого етапу визначається комплекс параметрів звуку, але ці параметри не порівнюються з параметрами, які знаходяться в пам'яті людини. На другому етапі виконується первинне розпізнавання звуків мови та порівняння їх за параметрами, що знаходяться в пам'яті. На третьому етапі звук уточнюється за змістом та зв'язком з іншими звуками мови.

Розрізнюють дві групи методів аналізу фонем. Першу групу складають методи, за яких аналізуються спектральні, часові та спектрально-часові характеристики фонем, формантні частоти, часові інтервали між переходами мовного сигналу через нуль.

Методи другої групи аналізу фонем засновані на таких ознаках, як дзвінкість-глухість, шумність, тривалість та інтенсивність. Методи першої та другої груп пов'язані між собою, тому можуть використовуватися конкретні технічні рішення, де для аналізу фонем використовують ознаки першої та другої груп.

Встановлено, що за спектром погано розпізнаються приголосні фонеми. Для їх розпізнавання необхідно використовувати спектрально-часові характеристики, крутість зростання та спадання рівнів сигналу, тривалість фонем. Окрім спектрально-часових обвідних сигналів аналізують похідні обвідних, що дозволяє визначити не тільки місцеположення та рівень формант, але й швидкість зростання та спадання обвідних спектра. Введено поняття субформант – спектральних максимумів похідної за часом від часової обвідної у вузькій смузі частот мовного сигналу. Субформанти оцінюються місцерозташуванням та рівнями спектральних максимумів у вузькій смузі частот часової обвідної мовного сигналу. Субформанти розміщені в діапазоні частот 16…76 Гц і дуже важливі при визначенні приголосних фонем. Існує також поняття супроводжуючих формант, які являють собою спектральні максимуми сумарної часової обвідної сигналу в діапазоні частот. Розпізнавання фонем проводиться за формантами, субформантами та супроводжуючими формантами.

На рис. 5.20 наведена схема автоматичного виділення фонем.

Мовний сигнал після компресора, який на рис. 5.20 не показаний, поступає на смугові фільтри (СФ). Перші вісім каналів (1…8) називаються ***формантними*** (основними), а два останніх (9, 10) – ***шумовими***. Шумові канали обумовлюють визначення шумових фонем, а нульовий канал (0) призначений для розпізнавання голосних та сонорних фонем [6].

З виходів смугових фільтрів сигнали через детектори Д подаються на фільтри нижніх частот ФНЧ зі смугою 30 Гц. Сигнали сумарного каналу $\sum$, що виділені в смузі 100…6000 Гц, надходять на блок виділення супровідних формант СФ0.

У каналах 2…7 проводять виділення субформант за допомогою субформантних блоків СФ2…СФ7. Сигнал у кожному з цих каналів після ФНЧ диференціюють, а потім за допомогою напівперіодних детекторів різної полярності розділяють зростання та спадання часової обвідної сигналу. Чим крутіше спадання або піднімання часової обвідної, тим більші амплітуди диференційованих сигналів. Диференціювання обвідної, а також розділення зростання та спадання часової обвідної, проводиться і в сумарному каналі.

Кожний із 12 сигналів, отриманих на виході блоків СФ2…СФ7, ділять на дві смуги: 16…46 та 46…76 Гц. Отримані 24 сигнали детектують та згладжують за допомогою фільтрів нижніх частот.

Розпізнавання фонем в одному з варіантів вокодера здійснюється за 48 параметрами, кожний з яких дискретизується за величиною на три значення (0, 1, 2). Для передачі частоти основного тону мови використовують 16 дискретних

значень в діапазоні 76…344 Гц. Для передачі кожного з 48 параметрів необхідно по 3 біти, а для передачі ОТ достатньо 4 біти. Якщо сигнали дискретизуються з інтервалами 1/15 с, то загальна швидкість передачі складатиме

$$(48 \times 3 + 4) \times 15 = 1220 \text{ біт/с}.$$

Така швидкість передачі забезпечується при використанні формантних вокодерів. Для зниження необхідної швидкості передачі в аналізаторі фонемного вокодера проводиться ідентифікація до 1000 варіантів фонем (близько 150 позиційних варіантів, кожний з яких в середньому має по шість варіантів). Для передачі 1000 варіантів фонем треба 10 біт, тому швидкість передачі фонем буде: 10х15=150 біт/с. З урахуванням 60 біт, які потрібні для передачі основного тону, загальна необхідна швидкість передачі буде дорівнювати 210 біт/с. Природньо, що в пам'яті синтезатора фонемного вокодера треба зберігати вказані 1000 варіантів фонем.

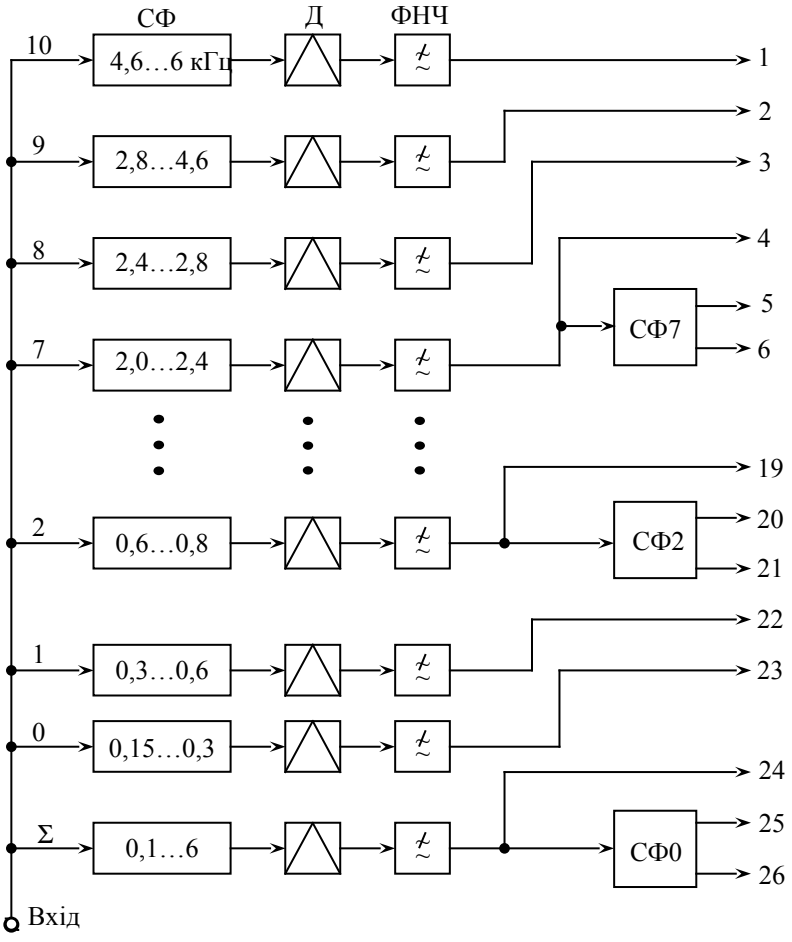


Рисунок 5.20 – Автоматичне виділення фонем:
СФ – смугові фільтри; Д – детектори; ФНЧ – фільтри нижніх частот; СФ2…СФ7 – субформантні блоки; СФ0 – блок супроводідних формант

Якщо не передавати варіанти фонем, а тільки класифікувати кожну фонему, то при інтервалі дискретизації 1/15 с, необхідна швидкість передачі складе всього 90 біт/с, тому що для передачі інформації про фонему достатньо 6 біт.

Коли виділити ще 60 біт/с для передачі основного тону мови, то для організації одного телефонного каналу за допомогою фонемного вокодера буде потрібна швидкість цифрового потоку 150 біт/с.

Для розпізнавання фонем використовують кореляційний метод. Суть його в тому, що спектр мовного сигналу ділиться на ряд смуг і в кожній смузі проводиться перемноження спектра даної фонеми на спектри зразкових фонем. Зразок, який забезпечує найбільший добуток у даному відрізку часу, відповідає вимовляній фонемі.

Розроблена бінарна система виділення фонем. При реалізації цієї системи фонеми окремими блоками діляться на два класи. Схема визначення фонем показана на рис. 5.21 [6].

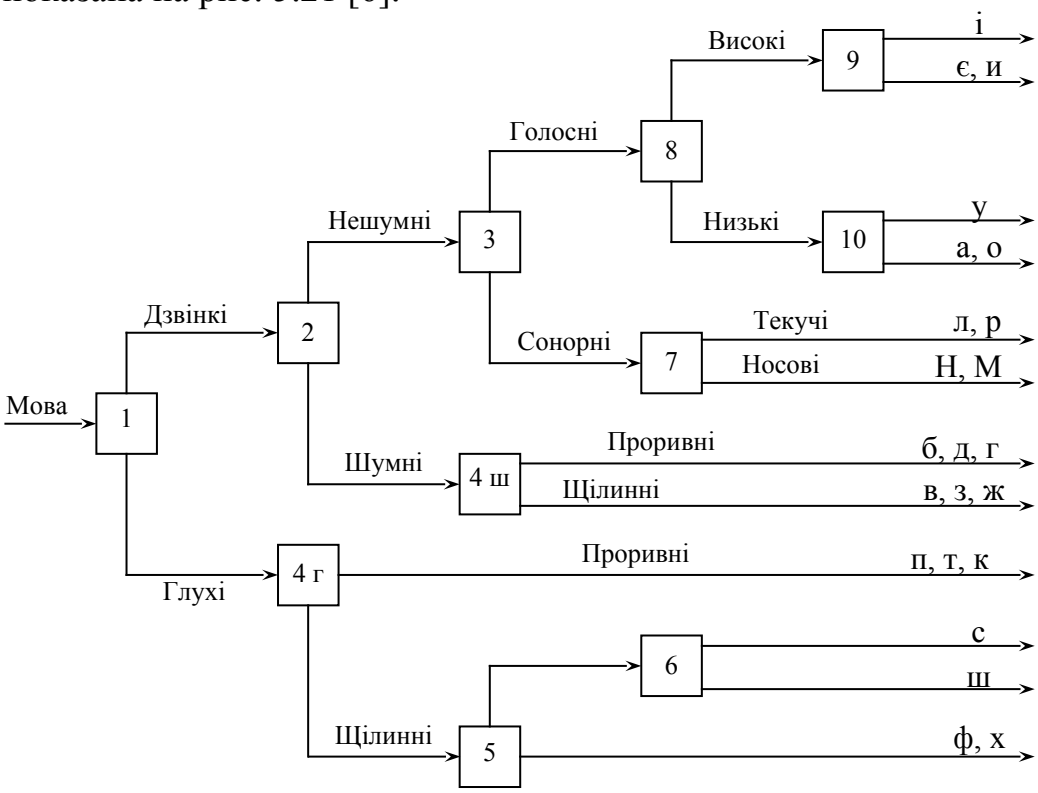


Рисунок 5.21 – Схема бінарного визначника фонем

Блок 1 за ознакою наявності або відсутності основного тону розподіляє фонеми на дзвінкі та глухі. При вимові дзвінких фонем існує основний тон. Відсутність основного тону свідчить про те, що фонема глуха. Блок 2 ділить шумові та нешумові фонеми, аналізує спектральні складові в області першої форманти та в більш високочастотній області. Шумова фонема має більший рівень в високочастотній області. Блок 3 за рівнем спектральних складових в області низьких та високих частот ділить голосні та сонорні. Голосні мають великий рівень в області низьких частот. Таким же чином блок 8 розподіляє високі та низькі голосні.

Інші блоки схеми використовують як розглянуті вище, так і інші ознаки. Наприклад, у блоці 6 фонеми «с» та «ш» розрізняються за частотою переходу сигналу через нуль. Число таких переходів при вимові фонеми «с» більше. Фо-

неми «л» і «р» розрізнюють за часовою обвідною сигналу. Для фонеми «р» характерні перерви часової обвідної, а обвідна фонеми «л» більш плавна.

Для розпізнавання фонем також використовують таблиці, в яких містяться ознаки фонем за трійковою системою (+,–,0). Знак «+» відповідає наявності ознаки, а знак «–» – відсутності ознаки. Знак «0» означає невизначеність. Фонеми можуть розпізнаватись за типовими спектрами, зразки яких є у пам'яті.

Для розпізнавання фонем можуть використовувати і ЕОМ. При цьому використовують методи, розглянуті раніше, та додаткові ознаки фонем, які містяться в структурі сигналу.

При різних вимовах розпізнавання фонем не може бути стовідсотковим. Навіть за наявності змістового зв'язку не завжди можна розпізнати передану мову, якщо у абонента погана дикція або є завади.

Синтез фонем можна здійснити по заздалегідь записаних натуральних звуках, за правилами мовоутворення і за допомогою електричного еквіваленту мовного тракту. В пам'яті синтезатора зберігаються різні варіанти фонем і, використовуючи їх кодові позначення, вони вибираються з пам'яті і по них відновлюються слова. Якщо взяти число варіантів фонем не менше 150, то стики фонем майже не сприймаються слухом, коли ці фонеми записані в пам'ять при звучанні одного голосу. Така мова не передає індивідуальні відтінки голосу, оскільки фонеми заздалегідь записані в пам'яті синтезатора.

При синтезі мови за правилами враховують наголос, інтонацію, тривалість звучання фонем. Розпізнавання фонем при цьому покращується, звучання мови стає більш натуральним.

Розроблені синтезатори, в яких є електричний еквівалент мовного тракту. Такий еквівалент складається з контурів. Частоти резонансів контурів можуть змінюватись при надходженні сигналів відповідних фонем. До контурів підключаються сигнали генератора пилкоподібних імпульсів або генератора шуму. Існують пристрої, які дозволяють знаходити помилки при аналізі та синтезі мови за допомогою правил вимови.

З мовного сигналу складно виділяти окремі фонеми, тому розроблені пристрої, які розпізнають типові комбінації звуків. В пам'яті повинні знаходитись ці комбінації і вибиратись в міру необхідності. Існують словесні та командні розпізнювачі, які визначають слова за спектрально-часовим розподіленням, зміною рівнів звучання, місцеположенням та зміною формант. Команди та склади вибираються з пам'яті за кодовими сигналами, які надходять по каналу зв'язку. Слова також можуть розпізнаватись за числом імпульсів та густиною нульових переходів. На цій основі розроблені пристрої, які розрізнюють до 1000 слів, які за кодовими сигналами вибираються з пам'яті. Такі пристрої можуть конкурувати з фонемними методами передачі, якщо використовується обмежене число слів або команд. Один з таких пристроїв, який автоматично розпізнає 35 слів, використовується в авіації.

Не виключено застосування автоматичних пристроїв, які розпізнають велику кількість слів. Середня тривалість вимови одного слова близько 1 с. Якщо закодувати 16000 слів, то необхідна пропускна здатність каналу зв'язку буде дорівнювати 14 біт/с.

З метою зменшення пам'яті доцільно кодувати різні комбінації звуків мови. Прийнятною є передача діад, які являють собою комбінації кінцевої половини одного звуку та початкової половини наступного. Кількість діад у різних мовах складає 600…1200. Кожна діада може застосовуватись з різною тривалістю, інтонацією та наголосом. У пам'яті треба зберігати всі варіанти діад.

Можна при передачі застосовувати триплети (кінцева половина першого звуку, другий звук та перша половина третього звуку). Число таких сегментів у російській мові не більше 4000. Ці сегменти (триплети) доцільно сполучати посередині голосних звуків. У цьому випадку з'єднання менше сприймається слухом. В даному випадку при одній і тій же пропускній здатності каналу зв'язку, необхідно мати менший об'єм пам'яті аналізатора та синтезатора, ніж при передачі слів.

5.7. Застосування вокодерів

У попередніх розділах розглянуті основні методи обробки звукових сигналів. Ці методи розглядались з точки зору передачі мовних сигналів і сигналів звукового мовлення. Розглянуті методи аналізу та синтезу мови знаходять застосування не тільки в галузі зв'язку, але й в інших областях техніки, а саме: в інформаційних мовних автоматах, при проведенні медичних обстежень, при зв'язку з водолазами, керуванні голосом технологічними процесами, введенні та виведенні мови з банків даних ЕОМ, ідентифікації особи, навчанні правильній вимові іноземців та осіб з дефектами мови, читанні книжок незрячими, при розпізнаванні мови автоматичною друкарською машинкою. Вказані області плідно розвиваються, тому важко розглянути всі особливості пристроїв, які аналізують та синтезують мову.

Системи обміну між людиною та машиною можуть бути з мовною відповіддю, з розпізнанням мови та розпізнанням абонента. Системи з мовною відповіддю являють собою системи одностороннього зв'язку, від машини до людини. Системи розпізнавання мови призначені для розпізнавання голосів різних людей. Особливий клас складають системи розпізнавання абонента. Ці системи являють собою два підкласи: верифікації та ідентифікації абонента. Системи верифікації встановлюють особу абонента за голосом, а системи ідентифікації можуть за голосом виділяти особу абонента з множини інших абонентів.

Система з мовною відповіддю отримує відповідь на питання від системи обробки інформації. Прикладом такої системи може бути автоматична телефонна служба, яка знаходить невірно набраний номер, і дає абоненту необхідні вказівки у вигляді мовної відповіді. Словник таких систем складається з обмеженої кількості слів та формується занесенням окремих слів реального мовного сигналу в пам'ять. Подання та зберігання словника повинно бути таким, щоб існувала можливість доступу до будь-якого слова. Також необхідно забезпечити спосіб формування мовних відповідей з дотриманням певних правил [6].

Використовують системи для видачі мовних команд монтажнику при виконанні міжблочних з'єднань апаратури. Вказівки складаються з коротких фраз,

які містять інформацію про колір, довжину проводу та місце його підключення. Інструкція складається з невеликого словника близько 100 слів. Після виконання команди з допомогою ножної педалі, монтажник надсилає запит на отримання наступної команди. Інструкція з монтажу у вигляді карток легко замінюється іншою [6].

На базі ЕОМ будуються мовні автомати, в пам'яті яких зберігаються сигнали керування, правила синтаксису, словник вимови слів, включаючи транскрипцію та наголоси. ЕОМ готує відповідь на запит абонента у вигляді тексту, що обробляється програмою синтезу. При створенні складної мови імітується динаміка мовного тракту з урахуванням акустичних особливостей мовоутворення. Якість такої мови відповідає натуральній. Такі автомати можна використовувати в інформаційно-довідковій службі, при читанні тексту незрячими та для навчання іноземців правильній вимові.

Розпізнавання мови важливе при створенні машин, керуючих технологічними процесами. Розроблені фонетичні друкарські машинки, які розпізнають мову і після корекції видають на друкування. Для розпізнавання елементів мови використовують методи, розглянуті в розд. 5.6. У Франції створений пристрій ідентифікації абонента за ім'ям, вимовленим у мікрофон, та здійснюючий автоматичне набирання номера потрібного телефону.

При зверненні абонента по телефону до банку даних, де утримується інформація про рахунки та інші довідки, доступ до яких обмежений, необхідно встановити справжність особи абонента. Задача верифікації вирішується порівнянням парольної фрази з еталоном, який є у пам'яті ЕОМ. Якщо при верифікації за паролем справжність голосу абонента встановлена, то потрібні йому відомості видаються по телефону. Після встановлення справжності голосу абонента ЕОМ може виконувати такі прості доручення як переказ грошей на інший рахунок, сплату вартості послуг та товарів, придбання квитків, пошук необхідних відомостей в базі даних і т.д. Причому, виконаня вказаних операцій супроводжується мовним діалогом людини та машини, яка за необхідності уточнює деякі деталі при виконанні доручень.

Верифікація голосу абонента проводиться за рядом параметрів парольної фрази, а саме: за середнім спектром тональних та шумових ділянок мови, за середнім значенням рівнів обвідних мовного сигналу та його похідних, за зміною частоти основного тону та її похідної. Оцінюються тривалості та спектри слів парольної фрази. Для верифікації можуть використовуватись й інші параметри мовного сигналу (форманти, антиформанти, субформанти).

При створенні еталону, що зберігається в пам'яті ЕОМ, абонент повинен декілька разів повторити парольну фразу по телефону. Чим більше параметрів піддається аналізу, тим точніше проводиться верифікація. Перевірка одного з алгоритмів верифікації показала, що помилка верифікації 10 абонентів за допомогою ЕОМ склала 1%. Помилки слухачів при верифікації на слух голосів абонентів були близько 4%. Спроби імітації голосів абонентів іншими особами були безрезультатними [19].

Задачі верифікації та ідентифікації багато в чому співпадають з точки зору обробки сигналів. Задача ідентифікації може виникнути в слідчій прак-

тиці, коли треба виділити голос одного абонента з більшості інших. Вирішується це питання порівнянням однієї або декількох фраз. Алгоритм роботи ЕОМ при ідентифікації особи лишається таким самим, як і при верифікації.

Методи аналізу звукових сигналів використовують при проведенні медичних обстежень. Аналіз коливань голосових зв'язок дозволяє виявити захворювання на ранніх стадіях, коли звичайні методи медичного обстеження малоефективні. Це завдання успішно вирішується за допомогою ЕОМ, яка має пристрій вводу акустичних сигналів та відповідне програмне забезпечення [19].

Широко використовують вокодери в системах телефонного зв'язку. Аналогові вокодери останнім часом не використовують через високу вартість обладнання. Переважно вокодери застосовують там, де канали зв'язку мають малу ємність або їх експлуатація коштує дорого.

Вокодери використовують при цифровій передачі телефонних сигналів по каналах сотових систем радіозв'язку, при передачі сигналів в КХ діапазоні, в супутникових системах передачі. Фонемні вокодери через низьку якість показників застосовують, в основному, в системах службового та комерційного зв'язку. Смугові та формантні вокодери використовують при передачі телефонних сигналів у мережах загального використання. При швидкості передачі 1200 біт/с найкращим за розбірливістю та натуральністю звучання сьогодні є формантний вокодер.

Вокодери можна порівняти за складністю технічної реалізації, якщо прийняти складність дельта-модулятора за одиницю. Тоді складність реалізації смугового вокодера складає близько 200 одиниць, формантного – 500 та фонемного – 1000 одиниць. Прогрес в області схемотехніки веде до здешевлення складних спеціалізованих мікросхем на базі яких будуються вокодери, тому складність побудови схеми не буде визначальною при виборі варіанту вокодера.

Вокодери застосовують для глибоководного телефонного зв'язку з водолазами. Водолази дихають сумішшю з підвищеним тиском. За рахунок цього спектр мови водолазів зміщується вверх за частотою, і мова стає нерозбірливою. В цих вокодерах частоти фільтрів аналізаторів вибирають в 1,5…2,0 рази вищими, ніж у звичайних вокодерах, а фільтри синтезатора розміщують в смугах частот натуральної мови [6].

Коли окремі ділянки каналу зв'язку обладнані вокодерами різних типів, то при переприйомі за низькою частотою різко погіршується розбірливість і натуральність звучання мовних сигналів. Причиною погіршення розбірливості та звучання сигналів є повторні вокодерні перетворення, оскільки переприйом за низькою частотою складається з синтезу та аналізу сигналів. Якість синтезованої мови значною мірою визначається основним тоном, а виділення основного тону з синтезованої мови супроводжується помилками, що підтверджено експериментально. За осцилограмами стандартної фрази оцінювали точність виділення основного тону фрази, прочитаної трьома дикторами-чоловіками. Точність виділення основного тону з натуральної мови була рівною 92%, а з синтезованої мови – 77%, тобто число помилок в другому випадку збільшилась майже втричі. При переприйомі за низькою частотою збільшується приблизно

вдвічі число помилок розпізнавань фонем. Основною причиною збільшення числа помилок також є вокодерні перетворення. Помилки з'являються і при зміні часової обвідної спектра мовних сигналів, якщо переприйом проводиться за низькою частотою.

Кількість помилок суттєво зменшується, якщо в місцях стиковок ділянок каналів зв'язку, обладнаних різними типами вокодерів, включені спеціальні узгоджувальні пристрої. Ці пристрої забезпечують перетворення сигнал-параметрів одного типу вокодера в сигнал-параметри вокодера іншого типу. Перетворення сигнал-параметрів проводиться без синтезу та повторного аналізу мовних сигналів. Використовуючи часові обвідні спектральних каналів вокодера першого типу, обчислюють часові обвідні спектральних каналів вокодера другого типу. При переході від більш високої швидкості передачі до більш низької знижують точність передачі часової обвідної спектра або зменшують число передаваних коефіцієнтів, а при передачі в зворотньому напрямку заповнюють вільні інфомаційні позиції за допомогою інтерполяції.

Сучасна техніка телефонного зв'язку дозволяє організовувати групові наради по телефону з участю трьох і більше абонентів. Учасники наради можуть одночасно говорити і слухати виступ кожного учасника (нарадний циркулярний зв'язок).

При використанні вокодерів режим нарадного циркулярного зв'язку організувати неможливо, тому що при об'єднані сигнал-параметрів двох і більше вокодерів погіршується синтез мови. Можна в комутаторі змішувати сигнали, які пройшли через синтезатор, тобто змішування сигналів абонентів буде здійснюватись за низькою частотою. Далі змішані сигнали можна подати на аналізатори. В цьому випадку зростають завади за рахунок переприйому за низькою частотою, але абоненти можуть слухати і говорити одночасно. Якщо виникає необхідність проведення подібних нарад, то учасники повинні враховувати, що натуральність звучання і розбірливість мови будуть гірші, ніж при звичайному двосторонньому вокодерному зв'язку.

Є цікавий приклад застосування вокодерів у системі мовного спілкування людини з машиною. Абонент по телефону замовляє авіаквиток. ЕОМ за парольною фразою перевіряє наявність грошей на поточному рахунку абонента, вибирає підходящий рейс, перевіряє наявність квітків, уточнює бажану вартість (класність) квитка, вказує в який час буде передано квиток абоненту і подано рахунок. При виконанні кожної з операцій проходить мовний діалог людини і машини.

Контрольні питання

1. В яких характеристиках міститься основна інформація про розбірливість мови?
2. Які параметри виділяють в параметричних вокодерах?
3. Що використовують при визначенні величини основного тону мови?
4. Які основні ознаки відміни дзвінких і глухих звуків мови?
5. Які пристрої називають напіввокодерами?

6. Які сигнали використовують при синтезі дзвінких і глухих звуків мови в смуговому вокодері?

7. В яких межах вибирають число частотних каналів в смугових вокодерах?

8. Яка величина смуги пропускання фільтрів нижніх частот смугового вокодера?

9. Скільки формант використовують у формантних вокодерах?

10. Як вимірюють формантну частоту?

11. Як можна визначити формантні частоти за допомогою фільтрів?

12. Які вокодери називають фазовими?

13. В чому полягає принцип роботи гармонічного вокодера?

14. Що являють собою сигнал-параметри кореляційного вокодера, які передають по каналу зв’язку?

15. Які перетворення називають гомоморфними?

16. Дайте визначення кепстра потужності.

17. Чому відповідають максимуми кепстра потужності звуків мови?

18. Які вокодери називають ліпредерами?

19. Для чого використовується кодова книжка в алгоритмі CELP?

20. Які функції виконує фільтр зважування в алгоритмі CELP?

21. Яка роль пост-фільтра в декодері CELP?

22. Якими технологіями забезпечується передача мовних сигналів з змінною швидкістю?

23. Дайте визначення субформанти.

24. Скільки біт необхідно виділити для передачі по каналу зв’язку номера фонеми?

25. Дайте визначення діад і триплетів.

26. Дайте визначення верифікації та ідентифікації абонента.

ЗАКІНЧЕННЯ

Розглянуті основні методи стиснення сигналів не вичерпують всіх особливостей побудови конкретних технічних реалізацій. Досить детально розглянуті методи стиснення сигналів у часовій області, які широко використовують у сучасній техніці. Схематично подані матеріали зі стиснення сигналів у частотній області, тому що в літературі описані тільки деякі аспекти побудови апаратури, а схеми кодерів і декодерів є «ноу-хау» фірм-розробників.

Розвитку техніки стиснення звукових сигналів заважає відсутність даних про психоакустичні характеристики вуха людини в різних ділянках спектра та залежність цих характеристик від рівня маскувальних сигналів. У літературі є тільки загальні властивості взаємного маскування сигналів. Сучасній рівень техніки стиснення дозволяє при допустимому ускладненні апаратури враховувати особливості психоакустичних характеристик в різних ділянках спектра, якби такі характеристики були достовірно відомі. В даний час враховують в апаратурі тільки мізерні дані про взаємне маскування сигналів.

Література

1. Орищенко В.И. Сжатие данных в системах сбора и передачи информации/ В.И. Орищенко, В.Г. Санников, В.А. Свириденко. – М.: Радио и связь, 1985. – 184 с.

2. Назаров М.В. Методы цифровой обработки и передачи речевых сигналов/ М.В. Назаров, Ю.Н. Прохоров. – М.: Радио и связь, 1985. – 176 с.

3. Сапожков М.А. Электроакустика: учебник для вузов/ М.А. Сапожков. – М.: Связь, 1978. – 272 с.

4. Римский-Корсаков А.В. Электроакустика/ А.В. Римский-Корсаков. – М.: Связь, 1973. – 272 с.

5. Цвикер Э. Ухо как приемник информации/ Э. Цвикер, Р.Фельдкеллер; пер. с немецкого под ред. Б.Г. Белкина. – М.: Связь, 1971. – 255 с.

6. Сапожков М.А. Вокодерная связь/ М.А. Сапожков, В.Г. Михайлов. – М.: Радио и связь, 1983. – 248 с.

7. Справочник по акустике / Иофе В.К., Корольков В.Г., Сапожков М.А.; под общ. ред. М.А.Сапожкова. – М.: Связь, 1979. – 312 с.

8. Справочник по радиовещанию / Выходец А.В., Захарин В.М., Рудый Е.М., Денисов В.И.; под общ. ред. А.В. Выходца. – К.: Техника, 1981. – 264 с.

9. Дворецкий И.М. Цифровая передача сигналов звукового вещания/ И.М. Дворецкий, И.Н. Дриацкий. – М.: Радио и связь, 1987. – 192 с.

10. CCITT. Rec. J. 41. Final Rep. to the VIII – PA, Rep. R46, 1984. – P. 17–29.

11. Глухов А.А. Статистические исследования скважности сигналов программ центрального вещания// – Электросвязь. – 1972. – №6. – С. 4–11.

12. Колмогоров А.Н. Интерполирование и экстраполирование стационарных случайных последовательностей// А.Н. Колмогоров. Изв. АН СССР, 1941. – Т.5. – С. 3–14. – (серия математика).

13. Wiener N. The interpolation, extrapolation and smoothing of stationary time series// N. Wiener. – Wiley: New York, 1949.

14. Эшер Р.Б. Литература по адаптивным системам управления/ [Эшер Р.Б. и др.]// ТИИЭР. – 1976. – Т. 64. – № 8. – С. 126–143.

15. Шеннон К.Э. Математическая теория связи. В кн.: Работы по теории информации и кибернетике/ К.Э. Шеннон. – М.: ИЛ, 1963. – 830 с.

16. Емельянов Г.А. Передача дискретной информации: учебник для вузов/ Г.А. Емельянов, В.О. Шварцман. – М.: Радио и связь, 1982. –240 с.

17. Пилипчук И.Н. Адаптивная импульсно-кодовая модуляция/ И.Н. Пилипчук, В.П. Яковлев. – М.: Радио и связь, 1986. – 296 с.

18. Рабинер Л.Р. Цифровая обработка речевых сигналов/ Л.Р. Рабинер, Р.В. Шафер; пер. с англ.; под ред. М.В. Назарова и Ю.Н. Прохорова. – М.: Радио и связь, 1981.– 496 с.

19. Маркел Дж. Линейное предсказание речи/ Дж. Маркел, А.Х. Грэй; пер. с англ.; под ред. Ю.Н. Прохорова и В.С. Звездина. – М.: Связь, 1980. – 308 с.

20. Денисов В.И. О критериях оценки цифровой обработки сигналов звукового вещания/ В.И. Денисов, Е.М. Рудый // Помехоустойчивость и эффе-

ктивность систем передачи информации: матер. Республиканск. научн.-техн. конф., 3-5 октября 1986 г.: тезисы докл.,– Одесса, 1986. С. 43.

21. Рудый Е.М. Заметность искажений при ортогональном преобразовании звуковых сигналов/ Е.М. Рудый // Проблемы цифровой записи: матер. Республиканск. конф., 4-6 октября 1988 г.: тезисы докл., – Алма-Ата, 1988.–С. 17.

22. Денисов В.И. Шумы обработки в системах сжатия сигналов звукового вещания/ В.И. Денисов, Е.М. Рудый // Совершенствование технической базы, организации и планирования телевидения и радиовещания: матер. Третьей Всесоюзн. научн.–техн. конф., 16-18 мая 1990: тезисы докл. – Москва, 1990. С. 180.

23. Денисов В.И. Исследование алгоритмов обработки звуковых сигналов в цифровых системах передачи с АДИКМ/ В.И. Денисов, Е.М. Рудый // матер. Девятой Всесоюзн. конф. по теории кодирования и передачи информации, 18-20 октября 1988 г.: тезисы докл. – Одесса, 1988. С. 312–315.

24. Денисов В.И. Исследование эффективности двухполосной ЦСП с АДИКМ/ В.И. Денисов, Е.М. Рудый, Г.В. Рабинович // Спутниковая связь: реальность и перспективы: Труды международн. симпозиума. 1-5 октября 1990 г.: Одесса, 1990. С. 201–207.

25. Быков С.Ф. Цифровая телефония: [учеб. пособие для вузов] / С.Ф. Быков, В.И. Журавлёв, И.А. Шалимов // – М.: Радио и связь, 2003. – 144 с.

ЗМІСТ

Навчальне видання

Рудий Євген Михайлович

ТЕХНОЛОГІЇ ПЕРЕДАЧІ ДИСКРЕТНИХ ПОВІДОМЛЕНЬ

Модуль 1: Стиснення телефонних сигналів

Навчальний посібник

Редактор – Кодрул Л.А.

Комп’ютерне макетування – Кірдогло Т.В.