Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

факультет прикладних інформаційних технологій та електроінженерії
(повна назва факультету)
кафедра радіотехнічних систем
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(назва освітнього ступеня)

на тему: Метод кодування аудіосигналів для підвищення ефективності фазових вокодерів

Виконав(ла): студент(ка) 6 курсу, групи РАм-61
спеціальності
172 Електронні комунікації та радіотехніка
(шифр і назва спеціальності)

Степанов В.Я.
(підпис) (прізвище та ініціали)

Керівник Дедів І.Ю.
(підпис) (прізвище та ініціали)

Нормоконтроль Хвостівська Л.В.
(підпис) (прізвище та ініціали)

Завідувач кафедри Дунець В.Л.
(підпис) (прізвище та ініціали)

Рецензент Тимків П.О.
(підпис) (прізвище та ініціали)

Тернопіль
2024

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет прикладних інформаційних технологій та електроінженерії
(повна назва факультету)

Кафедра радіотехнічних систем
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Дунець В.Л.
(підпис) (прізвище та ініціали)
« » 2024 р.

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня магістр
(назва освітнього ступеня)

за спеціальністю 172 Електронні комунікації та радіотехніка
(шифр і назва спеціальності)

студенту Степанову Валентину Ярославовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Метод кодування аудіосигналів для підвищення ефективності фазових вокодерів

Керівник роботи Дедів Ірина Юріївна, к.т.н., доцент
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «29» *листопада* 2024 року № 4/7-1145

2. Термін подання студентом завершеної роботи 11 грудня 2024 року

3. Вихідні дані до роботи Технічне завдання, фазові вокодери

4. Зміст роботи (перелік питань, які потрібно розробити)
1. Аналітична частина
2. Основна частина
3. Науково-дослідна частина
4. Охорона праці та безпека в надзвичайних ситуаціях

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці та безпека в надзвичайних ситуаціях	Окіпний І.Б., зав. каф. МТ		
	Зелінський І.М., доц. каф. ПВ		

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз завдання на кваліфікаційну роботу	06.09.2024	
2	Написання розділу 1	04.10.2024	
3	Написання розділу 2	21.10.2024	
4	Написання розділу 3	04.11.2024	
5	Написання розділу 4	18.11.2024	
6	Попередній захист	06.12.2024	
7	Захист	26.12.2024	

Студент _____ Степанов Валентин Ярославович
(підпис) (прізвище та ініціали)

Керівник роботи _____ Дедів Ірина Юріївна
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Метод кодування аудіосигналів для підвищення ефективності фазових вокодерів // Кваліфікаційна робота магістра // Степанов Валентин Ярославович // ТНТУ ім. І.Пулюя, ФПТ // Тернопіль, 2024 // с. - 74, рис. - 19, дод. - 1, бібл. - 30.

Ключові слова: АУДІОСИГНАЛ, ФАЗОВИЙ ВОКОДЕР, ВИСОТА ТОНУ.

В роботі проведено розроблення методу кодування аудіосигналів для підвищення ефективності фазових вокодерів. Проаналізовано особливості стиснення ГС в різного типу вокодерах. Проведено порівняльний аналіз алгоритмів кодування ГС, що лежать в основі роботи різних типів вокодерів та показано ефективність застосування класу фазових вокодерів. Проведено реалізацію основних етапів обробки ГС для фазових фокодерів в середовищі Matlab. Як алгоритм опрацювання використано підходи, описані в працях Ларош і Долсона, що передбачає часове стиснення/розтягування аудіосигналу, а також зміну висоти голосу диктора в аудіофайлі. Проведено моделювання процесу стиснення голосового сигналу та наступного його розтягування і оцінено похибку, яка характеризує процес кодування/декодування сигналу в фазовому вокодері.

ANNOTATION

Audio coding method to improve the efficiency of phase vocoders // Master's qualification work // Stepanov Valtntyn // TNTU, FPT // Ternopil, 2024 // p. - 74, fig. - 19, appl. - 1, bibl. - 30.

Key words: AUDIO SIGNAL, PHASE VOCODER, PITCH.

The paper develops a method for encoding audio signals to improve the efficiency of phase vocoders. The characteristics and structural parameters of the voice are analyzed to identify those that can be used for effective compression of voice signals in different types of vocoders. A comparative analysis of voice signal encoding algorithms underlying the operation of different types of vocoders is performed and the effectiveness of the phase vocoder class is shown. The main stages of voice signal processing for phase vocoders are implemented in the Matlab environment. The processing algorithm uses the approaches described in the works of Laroche and Dolson, which involves temporal compression/stretching of the audio signal, as well as changing the pitch of the speaker's voice in the audio file. The process of compressing the voice signal and its subsequent stretching is simulated and the error that characterizes the process of encoding/decoding the signal in the phase vocoder is estimated.

ЗМІСТ

ВСТУП

Актуальність теми. Фазовий вокодер — це тип алгоритму, що призначений для вокодера, який може інтерполювати інформацію, наявну в частотній і часовій областях аудіосигналів, використовуючи інформацію про фазу, отриману з частотного перетворення. Комп'ютерний алгоритм дозволяє модифікувати цифровий звуковий файл у частотній області (зазвичай розширення/стиснення за часом і зміна висоти тону).

В основі функціонування фазового вокодера лежить застосування короткочасного перетворення Фур'є (STFT), яке зазвичай кодується за допомогою швидких перетворень Фур'є. STFT перетворює представлення звуку в часовій області в частотно-часове представлення, дозволяючи модифікувати амплітуди або фази певних частотних компонентів звуку перед повторним синтезом представлення частотно-часової області в часову область за допомогою зворотного перетворення STFT.

Проблема перетворення сигналу за допомогою фазового вокодера пов'язана з проблемою того, що всі модифікації, які виконуються в поданні STFT, повинні зберігати відповідну кореляцію між сусідніми частотними елементами (вертикальна когерентність) і часовими межами (горизонтальна когерентність). За винятком надзвичайно простих синтетичних звуків, ці відповідні кореляції можуть бути збережені лише приблизно, і з моменту винаходу фазового вокодера дослідження в основному були спрямовані на пошук алгоритмів, які зберегли б вертикальну та горизонтальну узгодженість подання STFT після модифікації.

Таким чином, питання удосконалення чи розроблення нових алгоритмів кодування аудіосигналів, зокрема і голосових, для побудови фазових вокодерів, залишається актуальним.

Мета. Створення методу кодування аудіосигналів для задачі підвищення ефективності фазових вокодерів. **Задачі**:

- аналіз характеристик та структурних параметрів голосу для ефективного стиснення голосових сигналів (ГС) в різного типу вокодерах.;
- порівняльний аналіз алгоритмів перетворення аудіосигналів, зокрема ГСрізними вокодерами;
- аналіз особливостей роботи та недоліків фазових вокодерів
- запропонувати метод кодування аудіосигналів для фазових вокодерів і провсти його моделювання в середовищі Matlab.

Об'єкт дослідження: процес кодування аудіосигналів.

Предмет дослідження: метод перетворення аудіосигналів для підвищення ефективності фазових вокодерів.

Наукова новизна. Запропоновано метод та алгоритм реалізації процедури сповільнення/розтягування аудіосигналів та зміни висоти тону ГС для підвищення ефективності фазових вокодерів.

Практичне значення. Результати можна застосувати при створенні систем голосової аутентифікації користувачів на основі використання методів кодування ГС в фазових вокодерах.

РОЗДІЛ 1
АНАЛІТИЧНА ЧАСТИНА

1.1 Голосовий сигнал, як засіб обміну інформацією

Мовлення – це використання людського голосу як засобу спілкування. Розмовна мова поєднує різні звуки, щоб утворити одиниці, такі як слова, які належать до лексикону мови. Існує багато різних мовленнєвих актів, таких як інформування, оголошення, запит, переконання, скерування; дії можуть відрізнятися за різними аспектами, як-от вимова, інтонація, гучність і темп для передачі змістц. Люди також можуть ненавмисно повідомляти через мову аспекти свого соціального становища, такі як стать, вік, місце походження, фізіологічний і психічний стан, досвід.

Хоча зазвичай використовується для полегшення спілкування з іншими, люди також можуть використовувати мову без наміру спілкуватися. Проте мова може виражати емоції або бажання; люди іноді розмовляють самі з собою в актах, які є розвитком того, що деякі психологи стверджують, що це використання мовчазної мови у внутрішньому монолозі для оживлення та організації пізнання, іноді в миттєвому прийнятті подвійної персони як самозвернення до себе, як до іншої людини. Сольну мову можна використовувати для запам’ятовування чи перевірки запам’ятовування речей, а також у молитві чи медитації.

Дослідники вивчають багато різних аспектів мовлення: продукування мовлення та сприйняття звуків, що використовуються в мові, повторення мовлення, помилки мовлення, здатність відображати почуті вимовлені слова на вокалізацію, необхідну для їх відтворення, що відіграє ключову роль у розвитку дітей, їхнього словникового запасу та різних областей людського мозку, такі як зона Брока та зона Верніке, що лежать в основі мовлення. Мова є предметом вивчення лінгвістики, когнітивної науки, комунікації, психології, інформатики, логопедії, отоларингології та акустики. Мовлення порівнюється з письмовою

мовою, яка може відрізнятися своїм словниковим запасом, синтаксисом і фонетикою від усної мови.

Хоча еволюція характерних для людини мовленнєвих можливостей пов’язана із більш загальною проблемою походження мови, вона стала багато в чому окремою областю наукових досліджень. Тема широка, тому що мова не обов'язково є розмовна, вона однаково може бути письмовою. У цьому сенсі мовлення є необов’язковим, хоча це модальність мови за замовчуванням.

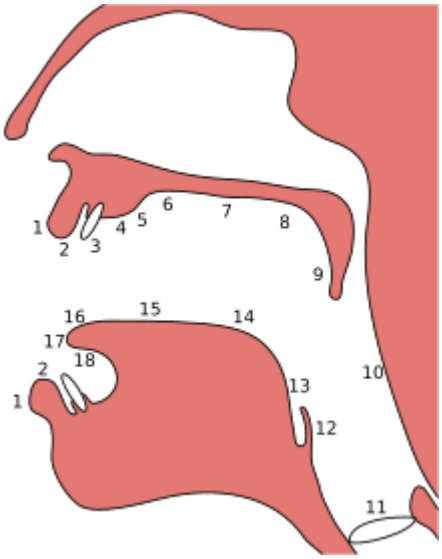

Рис. 1.1. Місця артикуляції (пасивні та активні): 1 - екзо-губне, 2 - ендо-губне, 3 - зубне, 4 - альвеолярне, 5 - заальвеолярне, 6 - переднебінне, 7 - піднебінне, 8 - веларне, 9 - увулярне, 10 - глоткове, 11 - глотальне, 12 - надгортанне, 13 - корінне, 14 - задньо-дорзальне, 15 - передньо-спинне, 16 - ламінальне, 17 - апікальне, 18 - підверхівкове

Генерування мовлення — це несвідомий багатоступінчастий процес, за допомогою якого думки перетворюються на усне висловлювання. Генерування включає підсвідомий вибір відповідних слів і відповідної форми цих слів із лексикону та морфології, а також організацію цих слів за допомогою синтаксису. Потім відновлюються фонетичні властивості слів, і речення артикулюється за допомогою артикуляцій, пов’язаних із цими фонетичними властивостями.

У лінгвістиці артикуляційна фонетика — це наука про те, як язик, губи, щелепа, голосові зв’язки та інші мовні органи використовуються для утворення

звуків. Звуки мови поділяються на категорії за способом артикуляції та місцем артикуляції. Місце артикуляції означає, де в шиї або роті звужується повітряний потік. Спосіб артикуляції стосується способу взаємодії мовних органів, наприклад, наскільки близько потік повітря обмежений, яка форма повітряного потоку використовується (наприклад, легенева, імплозивна, викидна), чи вібрують голосові зв'язки, і чи відкрита носова порожнина для потоку повітря. Це поняття в основному використовується для створення приголосних, але може використовуватися для голосних у таких якостях, як озвучення та назалізція. Для будь-якого місця артикуляції може бути кілька способів артикуляції, а отже, і кілька однорідних приголосних.

Нормальна людська мова є легеневою, створюється тиском з легенів, що створює фонацію в голосовій щілині в гортані, яка потім змінюється голосовим трактом і ротом на різні голосні та приголосні звуки. Однак люди можуть вимовляти слова без використання легенів і голосової щілини в гортанній мові, яка буває трьох типів: езофагеальна мова, фарингеальна мова та букальна мова

Мовлення є складною діяльністю, і, як наслідок, помилки є поширеними, особливо у дітей. Мовленнєві помилки часто використовуються при побудові моделей для продукування мови та засвоєння мови дитиною. Помилки мовлення, пов'язані з певними видами афазії, використовувалися для відображення певних компонентів мови в мозку та визначення зв'язку між різними аспектами продукування; наприклад, труднощі пацієнтів з експресивною афазією у створенні правильних дієслів минулого часу, але не неправильних, було використано, щоб продемонструвати, що регулярні відмінювані форми слова не зберігаються окремо в лексиконі, а утворюються з афіксації до базової форми.

Сприйняття мовлення стосується процесів, за допомогою яких люди можуть інтерпретувати та розуміти звуки, що використовуються в мові. Вивчення сприйняття мовлення тісно пов'язане з областями фонетики та фонології в лінгвістиці та когнітивній психології, сприйняття в психології. Дослідження сприйняття мовлення спрямоване на те, щоб зрозуміти, як слухачі розпізнають звуки мовлення та використовувати цю інформацію для розуміння розмовної мови. Дослідження сприйняття мовлення також знаходять

застосування у створенні комп’ютерних систем, які можуть розпізнавати мовлення, а також покращувати розпізнавання мовлення для слухачів із вадами слуху та мови.

Сприйняття мовлення є категоричним, оскільки люди поділяють звуки, які вони чують, на категорії, а не сприймають їх як спектр. Люди з більшою ймовірністю зможуть почути відмінності в звуках за межами категорій, ніж усередині них. Гарним прикладом цього є час початку голосу (VOT), один з аспектів фонетичного утворення приголосних звуків.

Під час повторення мовлення почута мова швидко перетворюється з сенсорного введення в рухові інструкції, необхідні для її негайної або відстроченої голосової імітації (у фонологічній пам’яті). Цей тип відображення відіграє ключову роль у розширенні словникового запасу дітей. Masur (1995) виявив, що те, наскільки часто діти повторюють нові слова порівняно з тими, які вони вже мають у своєму лексиконі, пов’язано з розміром їхнього лексикону пізніше, при цьому маленькі діти, які повторюють більше нових слів, мають більший лексикон пізніше. Повторення мовлення може допомогти полегшити засвоєння цього більшого лексикону.

1.2 Задача кодування мови

Кодування мови – це застосування стиснення даних до цифрових аудіосигналів, що містять мову. Кодування мовлення використовує оцінку специфічних для мовлення параметрів за допомогою методів обробки аудіосигналу для моделювання мовного сигналу в поєднанні з загальними алгоритмами стиснення даних для представлення отриманих змодельованих параметрів у компактному бітовому потоці.

Поширеними застосуваннями кодування мови є мобільна телефонія та IP голос (VoIP). Найбільш широко використовуваною технікою кодування мови в мобільній телефонії є лінійне передбачуване кодування (LPC), тоді як найбільш широко використовуваними в програмах VoIP є LPC і методи модифікованого дискретного косинусного перетворення (MDCT).

Техніки, які використовуються для кодування мовлення, подібні до тих, які використовуються для стиснення аудіоданих і аудіокодування, де оцінка психоакустики використовується для передачі лише даних, які стосуються слухової системи людини. Наприклад, при кодуванні смуги голосу передається лише інформація в діапазоні частот від 400 до 3500 Гц, але реконструйований сигнал зберігає належну розбірливість.

Кодування мови відрізняється від інших форм кодування звуку тим, що мова є простішим сигналом, ніж інші звукові сигнали, і доступна статистична інформація про властивості мови. Як наслідок, деяка звукова інформація, яка має відношення до загального звукового кодування, може бути непотрібною в контексті кодування мови. Кодування мови грунтується на збереженні розбірливості мови при використанні обмеженої кількості переданих даних. Крім того, більшість мовних програм вимагають низької затримки кодування, оскільки затримка заважає мовній взаємодії.

Кодери мови бувають двох класів:

- кодери хвиль
- кодери часової областф: PCM, ADPCM

Вокодери:

- лінійне кодування з прогнозуванням (LPC)
- кодування формант
- машинне навчання, тобто нейронний вокодер

Компандування зразків розглядається як форма кодування мови.

Алгоритми A-law і μ-law, що використовуються в цифровій телефонії G.711 PCM, можна розглядати як попередній метод кодування мови, вимагаючи лише 8 біт на вибірку, але даючи фактично 12 біт роздільної здатності. Логарифмічне компандування узгоджується зі сприйняттям людського слуху, оскільки шум низької амплітуди чується вздовж мовного сигналу низької амплітуди, але маскується шумом високої амплітуди. Незважаючи на те, що це призведе до неприйнятних спотворень в аудіосигналі, піковий характер мовних сигналів у поєднанні з простою частотною структурою мовлення як періодичної форми сигналу, що має одну основну частоту з випадковими доданими

шумовими сплесками, робить ці дуже прості алгоритми миттєвого стиснення прийнятними.

У той час було випробувано багато інших алгоритмів, переважно варіанти дельта-модуляції, але після ретельного розгляду розробники ранніх систем цифрової телефонії обрали алгоритми A-law/μ-law. На момент їх проектування їхнє зменшення пропускної здатності на 33% для дуже низької складності стало чудовим інженерним компромісом. Їх звукові характеристики залишаються прийнятними, і не було необхідності замінювати їх у стаціонарній телефонній мережі.

У 2008 році кодек G.711.1, який має масштабовану структуру, був стандартизований ITU-T. Вхідна частота дискретизації становить 16 кГц.

Сучасна компресія мови

Значна частина пізніших робіт зі стиснення мови була мотивована військовими дослідженнями цифрового зв’язку для захищених військових радіостанцій, де використовувалися дуже низькі швидкості передачі даних для ефективної роботи у ворожому радіосередовищі. У той же час у вигляді схем HBIC була доступна набагато більша потужність обробки, ніж для попередніх методів стиснення. Як результат, сучасні алгоритми стиснення мовлення можуть використовувати набагато складніші методи, ніж були доступні в 1960-х роках, щоб досягти набагато вищих коефіцієнтів стиснення.

Найбільш широко використовувані алгоритми кодування мови засновані на лінійному прогнозуючому кодуванні (LPC). Зокрема, найпоширенішою схемою кодування мови є метод LPC з кодовим збудженням (CELP), яке використовується, наприклад, у стандарті GSM. У CELP моделювання поділяється на два етапи: етап лінійного прогнозування, який моделює спектральну оболонку, і модель залишку лінійної моделі прогнозування на основі кодової книги. У CELP коефіцієнти лінійного передбачення (LPC) обчислюються та квантуються, як правило, як пари лінійних спектрів (LSP). На додаток до фактичного мовного кодування сигналу, часто необхідно використовувати канальне кодування для передачі, щоб уникнути втрат через помилки передачі. Щоб отримати найкращі загальні результати кодування,

методи кодування мовлення та кодування каналів вибираються парами, причому важливіші біти в потоці мовних даних захищені надійнішим канальним кодуванням.

Модифіковане дискретне косинусне перетворення (MDCT) використовується в техніці LD-MDCT, яка використовується у форматі AAC-LD, представленому в 1999 році. З тих пір MDCT отримав широке застосування в додатках Voice-over-IP (VoIP), таких як широкосмуговий аудіокодек G.729.1, представлений у 2006 році, Apple FaceTime (з використанням AAC-LD), представлений у 2010 році та Кодек CELT, представлений у 2011 році.

Opus — це безкоштовний аудіокодер. Він поєднує в собі орієнтований на мовлення алгоритм SILK на основі LPC і алгоритм CELT на основі MDCT з меншою затримкою, перемикаючись між ними або комбінуючи їх за потреби для максимальної ефективності. Він широко використовується для VoIP-дзвінків у WhatsApp.

Було продемонстровано ряд кодеків із ще нижчими бітрейтами. Codec2, який працює на швидкості до 450 біт/с, використовується в аматорському радіо. Наразі НАТО використовує MELPe, що забезпечує розбірливу мову на швидкості 600 біт/с і нижче. Також з’явилися нейронні вокодери: Lyra від Google забезпечує «чудову моторошну» якість при 3 кбіт/с. Satin від Microsoft також використовує машинне навчання, але використовує вищий настроюваний бітрейт і є широкосмуговим.

1.3 Метод лінійного передбачення

Лінійне кодування з прогнозуванням (LPC) виходить з припущення, що мовний сигнал виробляється зумером на кінці трубки (для дзвінких звуків), з періодичними додаванням шиплячих і зсимкових звуків (для глухих звуків, таких як шиплячі та вибухові). Незважаючи на те, що ця модель джерело-фільтр виглядає грубою, насправді є близьким наближенням до реальної генерації мови. Голосова щілина (проміжок між голосовими складками) створює дзижчання, яке характеризується своєю інтенсивністю (гучністю) і частотою (висотою).

Голосовий тракт (горло і рот) утворює трубку, яка характеризується своїми резонансами; ці резонанси породжують форманти, або посилені смуги частот у звукові. Шипіння та вибухання породжуються діями язика, губ і горла під час шиплячих і вибухових звуків.

LPC аналізує мовний сигнал, оцінюючи форманти, видаляючи їхні ефекти з мовного сигналу та оцінюючи інтенсивність і частоту складової, що залишилася. Процес видалення формант називається зворотною фільтрацією, а сигнал, що залишився після віднімання відфільтрованого змодельованого сигналу, називається залишком.

Числа, які описують інтенсивність і частоту вібрацій, форманти та залишковий сигнал, можна зберігати або передавати в інше місце. LPC синтезує мовний сигнал шляхом зворотного процесу: використовуючи параметри шуму та залишок для створення вихідного сигналу, використовуючи форманти для створення фільтра (який представляє трубку) і пропусчаючи сигнал джерела через фільтр, у результаті чого утворюється мова.

Оскільки мовні сигнали змінюються з часом, цей процес виконується на коротких фрагментах мовного сигналу, які називаються кадрами; зазвичай від 30 до 50 кадрів на секунду забезпечують розбірливу мову з хорошим стисненням.

Лінійне передбачення (оцінка сигналу) починається принаймні з 1940-х років, коли Норберт Вінер розробив математичну теорію для розрахунку найкращих фільтрів і предикторів для виявлення сигналів, прихованих у шумі. Невдовзі після того, як Клод Шенон створив загальну теорію кодування, роботу з прогностичного кодування виконали К. Чапін Катлер, Бернард М. Олівер і Генрі К. Гарісон. Пітер Еліас у 1955 році опублікував дві статті про прогнозне кодування сигналів.

Лінійні предиктори були застосовані для аналізу мови незалежно Фумітада Ітакура з Університету Нагоя та Шузо Сайто з Nippon Telegraph and Telephone у 1966 році та в 1967 році Бішну С. Атал, Манфред Р. Шредер і Джон Берг. Ітакура та Сайто описали статистичний підхід, заснований на оцінці максимальної правдоподібності; Атал і Шредер описали підхід адаптивного лінійного

передбачення; Бург окреслив підхід, заснований на принципі максимальної ентропії.

У 1969 році Ітакура та Сайто представили метод, заснований на частковій кореляції (PARCOR), Глен Каллер запропонував кодування мови в реальному часі, а Бішну С. Атал представив кодер мови LPC на щорічних зборах Акустичного товариства Америки. У 1971 році компанія Philco-Ford продемонструвала LPC у реальному часі з використанням 16-розрядного обладнання LPC. Технологію LPC розвинули Бішну Атал і Манфред Шредер протягом 1970-1980-х років. У 1978 році Атал і Вішванат та ін. з BBN розробив перший алгоритм LPC зі змінною швидкістю. Того ж року Атал і Манфред Р. Шредер з Bell Labs запропонували мовний кодек LPC під назвою адаптивне прогностичне кодування, який використовував алгоритм психоакустичного кодування, використовуючи властивості маскування людського вуха. Пізніше це стало основою для техніки перцептивного кодування, яка використовується у форматі стиснення звуку MP3, представленому в 1993 році. Лінійне передбачення з кодовим збудженням (CELP) було розроблено Шредером і Аталом у 1985 році.

LPC є основою для технології голосу через IP (VoIP). У 1972 році Боб Кан з ARPA разом з Джимом Форгі з Лінкольнської лабораторії (LL) і Дейвом Уолденом з BBN Technologies розпочали перші розробки в області пакетного мовлення, що врешті призвело до технології голосу через IP. У 1973 році, згідно з неофіційною історією Лінкольнської лабораторії, перший LPC у реальному часі 2400 біт/с був реалізований Едом Хофштеттером. У 1974 році перший двосторонній пакетний мовний зв'язок LPC у режимі реального часу був здійснений через мережу ARPANET зі швидкістю 3500 біт/с між Куллер-Гаррісоном і лабораторією Лінкольна.

Представлення коефіцієнтів LPC

LPC часто використовується для передачі інформації про спектральну огинаючу, і є стійким до помилок передачі. Передача коефіцієнтів фільтра безпосередньо небажана, оскільки вони дуже чутливі до помилок. Іншими

словами, дуже мала помилка може спотворити весь спектр, або, що ще гірше, невелика помилка може зробити фільтр передбачення нестабільним.

Існують більш просунуті представлення, такі як логарифмічне відношення площі (LAR), розклад лінійних спектральних пар (LSP) і коефіцієнти відбиття. З них, особливо розкладання LSP набуло популярності, оскільки воно забезпечує стабільність предиктора, а спектральні помилки є локальними для малих відхилень коефіцієнта.

LPC є найбільш широко використовуваним методом кодування та синтезу мови. Зазвичай використовується для аналізу та повторного синтезу мовлення. Він використовується як форма стиснення голосу телефонними компаніями, наприклад, у стандарті GSM. Він також використовується для захищеного бездротового зв’язку, де голос має бути оцифрований, зашифрований і надісланий через вузький голосовий канал.

Синтез LPC може бути використаний для побудови вокодерів, де музичні інструменти використовуються як сигнал збудження для змінного в часі фільтра, оціненого з мови співака. Це є популярно в електронній музиці.

Передбачувачі LPC використовуються в аудіокодеках Shorten, MPEG-4 ALS, FLAC, SILK та інших аудіокодеках без втрат.

1.4 Вокодери та водери

Вокодер — це категорія кодування мови, яка аналізує та синтезує сигнал людського голосу для стиснення аудіоданих, мультиплексування, шифрування.

Вокодер був винайдений у 1938 році Гомером Дадлі в Bell Labs як засіб синтезу людської мови. Ця робота була реалізована у вигляді канального вокодера, який використовувався як голосовий кодек для телекомунікацій для кодування мови та збереження пропускної здатності при передачі.

За допомогою шифрування керуючих сигналів передачу голосу можна захистити від перехоплення. Його основне використання в цьому способі - для безпечного радіозв'язку. Перевагою цього методу є те, що вихідний сигнал не надсилається, лише результати роботи смугових фільтрів. Для повторного

синтезу версії оригінального спектру сигналу приймальний блок має бути налаштований у тій самій конфігурації фільтрів.

Вокодер також широко використовувався як електронний музичний інструмент. Декодерна частина вокодера, яка називається водером, може використовуватися незалежно для синтезу мови.

Голос складається зі звуків, що утворюються внаслідок періодичного відкриття та закриття голосової щілини голосовими зв'язками, що створює акустичну хвилю з багатьма гармоніками. Потім цей початковий звук фільтрується за допомогою рухів у носі, роті та горлі (складна резонансна система трубок), щоб створювати коливання гармонійного вмісту (форманти) у контрольований спосіб, створюючи широкий спектр звуків, які використовуються в мові. Існує ще один набір звуків, відомих як глухі та вибухові звуки, які створюються або змінюються різноманітними звукогенеруючими зрушеннями повітряного потоку, що відбуваються в голосовому тракті.

Вокодер аналізує мову, вимірюючи, як її характеристики спектрального розподілу енергії коливаються в часі. Результатом цього аналізу є набір паралельних у часі сигналів обвідної, кожен з яких представляє окрему амплітуду смуги частот мовлення користувача. Іншими словами, голосовий сигнал поділяється на кілька частотних діапазонів (чим більше це число, тим точніший аналіз), і рівень сигналу, присутній у кожній частотній смузі, що відбувається одночасно, виміряний повторювачем огинаючої, представляє спектральний розподіл енергії в часі. Цей набір сигналів амплітуди обвідної називається «модулятором». Щоб відтворити мовлення, вокодер змінює процес аналізу, змінно фільтруючи початковий широкосмуговий шум (який попеременно називають «джерелом» або «носієм»), пропускаючи його через набір смугових фільтрів, індивідуальні рівні амплітуди обвідної яких контролюються в режимі реального часу набором аналізованих сигналів амплітуди обвідної від модулятора.

Процес цифрового кодування передбачає періодичний аналіз кожного з багатосмугового набору амплітуд огинаючої фільтра модулятора. Результатом

цього аналізу є набір значень потоку цифрової ІКМ. Потім вихідний потік ІКМ кожного діапазону передається в декодер. Декодер подає ІКМ як керуючі сигнали до відповідних підсилювачів каналів вихідного фільтра.

Інформація про основну частоту початкового голосового сигналу (на відміну від його спектральної характеристики) відкидається; не було важливо зберегти це для початкового використання вокодера як допоміжного засобу шифрування. Саме цей дегуманний аспект процесу вокодування зробив його корисним для створення спеціальних голосових ефектів у популярній музиці та аудіорозвагах.

Замість покрокового відтворення форми хвилі вокодер надсилає лише параметри голосової моделі через канал зв'язку. Оскільки параметри змінюються повільно порівняно з вихідною формою сигналу мовлення, пропускну здатність, необхідну для передачі мовлення, можна зменшити.

Аналогові вокодери зазвичай аналізують вхідний сигнал, розділяючи сигнал на кілька налаштованих смуг або діапазонів частот. Щоб реконструювати сигнал, несучий сигнал надсилається через ряд цих налаштованих смугових фільтрів. У прикладі типового голосу робота носієм є шум або пилкоподібна форма сигналу.

Амплітуда модулятора для кожної з окремих смуг аналізу створює напругу, яка використовується для керування підсилювачами для кожної з відповідних смуг несучих. Результатом є те, що частотні компоненти модулюючого сигналу відображаються на несучому сигналі як дискретні зміни амплітуди в кожній із смуг частот.

Часто є глуха смуга або шиплячий канал. Це використовується для частот, які знаходяться за межами діапазонів аналізу для типового мовлення, але все ще важливі для мовлення. Прикладом є слова, які починаються з літер с, ф, ч або будь-якого іншого шиплячого звуку. Використання цієї смуги створює впізнавану мову, хоча й має дещо механічне звучання. Вокодери часто містять другу систему для генерації глухих звуків, використовуючи генератор шуму замість основної частоти.

В алгоритмі канального вокодера серед двох компонентів аналітичного сигналу врахування лише амплітудної складової та просте ігнорування фазової складової призводить до нечіткого голосу.

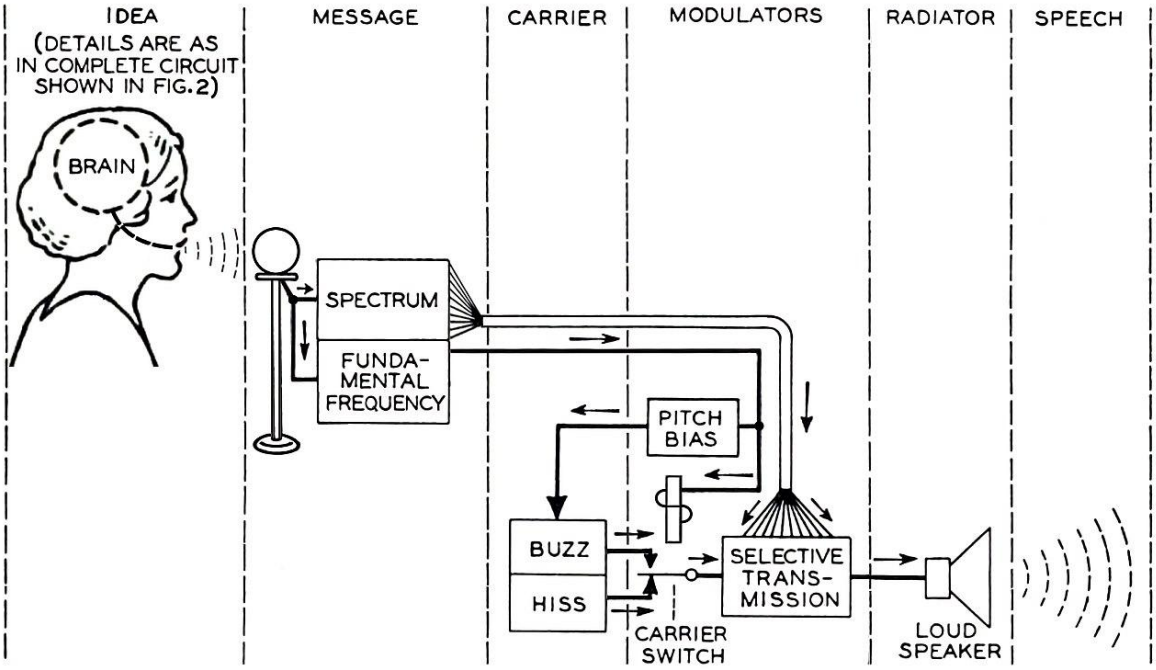

Рис. 1.2. Принципова схема вокодера Дадлі (1928р.)

Водер Дадлі складався з електронного генератора — джерела звуку високого тону — і генератора шуму для шипіння, 10-смугового резонаторного фільтра з підсилювачами зі змінним коефіцієнтом підсилення як вокального тракту та ручних контролерів, включаючи набір чутливих до тиску клавіш для керування фільтром і ножну педаль для керування висотою тону. Фільтри, керовані клавішами, перетворюють тон і шипіння на голосні, приголосні та флексії. Ця машина була складною для роботи, але досвідчений оператор міг створити впізнавану мову.

Вокодер Дадлі використовувався в системі SIGSALY, яку створили інженери Bell Labs у 1943 році. SIGSALY використовувався для зашифрованого голосового зв'язку під час Другої світової війни.

Сфери застосування:

Кінцеве обладнання для систем на основі цифрового мобільного радіо (DMR).

Цифрове кодування та шифрування голосу

Кохлеарні імплантати: для імітації ефектів кохлеарних імплантів використовується шумове та тональне вокодування.

Музичні та інші художні ефекти

Навіть при необхідності запису кількох частот і додаткових неозвучених звуків стиснення вокодерних систем вражає. Стандартні системи запису мовлення вловлюють частоти приблизно від 500 до 3400 Гц, де знаходиться більшість частот, що використовуються в мовленні. Роздільна здатність дискретизації зазвичай становить 8 або більше біт на роздільну здатність семплування для швидкості передачі даних у діапазоні 64 кбіт/с, але хороший вокодер може забезпечити досить хорошу імітацію голосу лише з 5 кбіт/с даних.

Голосові кодери ITU G.729, використовуються в багатьох телефонних мережах. G.729, зокрема, має кінцеву швидкість передачі даних 8 кбіт/с із чудовою якістю голосу. G.723 досягає трохи гіршої якості на швидкості передачі даних 5,3 і 6,4 кбіт/с. Багато систем голосових вокодерів використовують нижчу швидкість передачі даних, але нижче 5 кбіт/с якість голосу починає швидко падати.

Кілька систем вокодерів використовуються в системах шифрування NSA:

LPC-10, FIPS Pub 137, 2400 біт/с, який використовує лінійне кодування з прогнозуванням

CELP, 2400 і 4800 біт/с, федеральний стандарт 1016, використовується в STU-III

Безперервно змінна дельта-модуляція нахилу (CVSD), 16 кбіт/с, використовується в широкосмугових шифраторах, таких як KY-57.

Лінійне передбачення зі змішаним збудженням (MELP), MIL STD 3005, 2400 біт/с, використовується в майбутньому вузькосмуговому цифровому терміналі FNBDT, захищеному телефоні АНБ 21 століття.

Адаптивна диференційно-імпульсна кодова модуляція (ADPCM), раніше ITU-T G.721, 32 кбіт/с, що використовується в захищеній телефонії STE.

Сучасні вокодери, які сьогодні використовуються в комунікаційному обладнанні та пристроях зберігання голосу, базуються на таких алгоритмах:

Лінійне передбачення з алгебраїчним кодовим збудженням (ACELP 4,7–24 кбіт/с)

Багатодіапазонне збудження (AMBE 2000 біт/с – 9600 біт/с)

Синусоїдально-імпульсне представлення (SPR 600 біт/с – 4800 біт/с)

Надійна розширена інтерполяція сигналу низької складності (RALCWI 2050, 2400 і 2750 біт/с)

Трихвильове збуджене лінійне передбачення (TWELP 300–9600 біт/с)

Вокодери також зараз використовуються в психофізиці, лінгвістиці, обчислювальній нейронауці та дослідженні кохлеарних імплантів.

В кінці 70-х років більшість немузичних вокодерів було реалізовано з використанням лінійного передбачення, за допомогою якого спектральна обвідна цільового сигналу (форманта) оцінюється всеполюсним IIR-фільтром. У кодуванні з лінійним передбаченням всеполюсний фільтр замінює банк смугових фільтрів свого попередника та використовується в кодері для відбілювання сигналу (тобто згладжування спектру) і знову в декодері для повторного застосування спектральної форми.

Однією з переваг цього типу фільтрації є те, що розташування спектральних піків лінійного предиктора повністю визначається цільовим сигналом і може бути настільки точним, наскільки це дозволено періодом часу для фільтрації. Це відрізняється від вокодерів, реалізованих з використанням банків фільтрів фіксованої ширини, де розташування спектральних піків обмежене доступними фіксованими смугами частот. Фільтрація LP також має недоліки в тому, що сигнали з великою кількістю складових частот можуть перевищувати кількість частот, які можуть бути представлені фільтром лінійного передбачення. Це обмеження є основною причиною того, що кодування LP майже завжди використовується в тандемі з іншими методами в кодерах голосу з високим ступенем стиснення.

Вокодер з інтерполяцією форми сигналу (WI) був розроблений в AT&T Bell Laboratories приблизно в 1995 році, а згодом AT&T розробила нескладну версію для конкурсу захищених вокодерів Міністерства оборони. Помітні

вдосконалення кодера WI були зроблені в Каліфорнійському університеті в Санта-Барбарі.

1.5 Типи вокодерів та особливості їхнього функціонування

Вокодери можуть бути голосоелементними а також параметричними.

За допомогою вокодера вимовлені фонетичні елементи (такі як фонеми) розпізнаються під час передачі, і передаються лише їхні номери.

Після отримання ці елементи синтезуються відповідно до правил створення мови або витягуються з пам’яті пристрою.

Фонематичні вокодери використовуються в мовних апаратах для командних ліній зв'язку, голосового керування, інформаційно-довідкових служб.

По суті, такий вокодер автоматично розпізнає звукове зображення і не визначає параметри звуку.

Параметричні вокодери розрізняють два типи параметрів.

- обвідної спектра ГС (функція фільтра).

- джерела голосу (функція генератора), - ЧОТ, його зміни в часі, шумові сигнали;

Приймач синтезує мову на основі цих параметрів.

В ортогональному вокодері миттєвий спектральний контур поділяється на серію послідовностей відповідно до вибраної системи ортогональних функцій.

Розраховані коефіцієнти розкладу надсилаються до приймача.

Широко використовуються гармонічні вокодери з розкладанням у ряд Фур'є.

Оскільки CLP описує функцію фільтра, залишковий сигнал містить інформацію про генерування мови та надсилається до приймача (може бути стиснутий ADIKM, ADM або методами лінійного прогнозування нижчого порядку).

Якість аудіо вокодера є функцією бітрейту, пропускної здатності та затримки обробки.

Якщо вокодер призначений для Інтернет-телефонії, розробники продукту мають врахувати ці тісно пов’язані характеристики.

Вокодери спільно використовують канали зв'язку і часто перевантажені корпоративні мережі та Інтернет іншими потоками інформації, тому максимальна швидкість має бути чи понижчою, особливо в невеликих офісних додатках.

Більшість вокодерів сьогодні працюють з фіксованою швидкістю.

Для додатків, які передбачають одночасну передачу голосу та даних, можливе створення алгоритму компресії паузи.

Метод реалізації механізмів компресії пауз є важливим для покращення якості передачі звуку, але переваги стиснення паузи часто не використовуються.

Проблема полягає в тому, що коли є багато фонового шуму, важко відрізнити мову від шуму.

Генерація шуму повинна здійснюватися таким чином, щоб кодер і декодер були синхронізовані.

Це забезпечує плавні переходи між активними та неактивними сегментами аудіо.

1.6 Порівняння методів кодування голосу

Усі методи кодування цифрового аудіо умовно розділяються на 2 групи: кодери сигналу і джерела.

Схеми кодування звуку можна далі розділити на 3 групи.

Є й інші, але три виділені групи є найбільш вивченими.

Основним завданням і функцією цих методів є аналіз вхідного сигналу, видалення надмірності та відповідне кодування інформаційної частини сигналу.

Щоб уповільнити цифровий потік, нам потрібно розробити дедалі складніші методи зниження надмірності.

Рис. 1.3 проілюстровано найпоширеніші стратегії кодування.

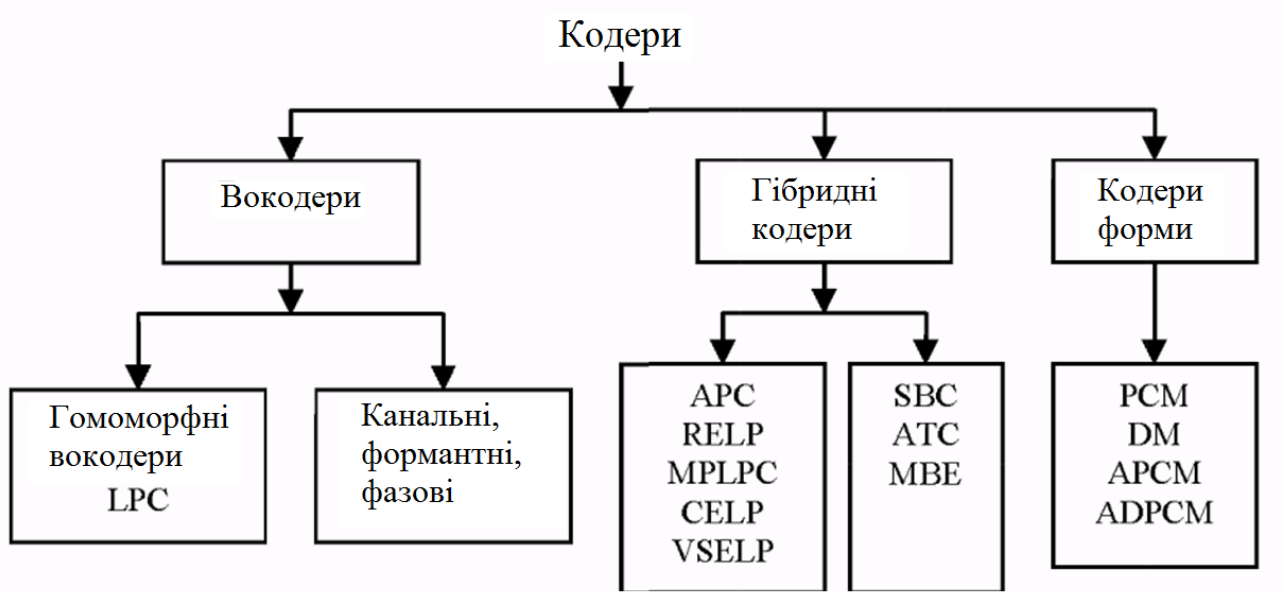

Рис. 1.3. Стратегії кодування ГС

Розглянемо параметри для порівняння схем кодування звуку.

Наступні параметри є найбільш важливими.

Швидкість — бажаний діапазон швидкості зв’язку.

Системи з нижчою швидкістю передачі даних вимагають меншої пропускної здатності, що призводить до кращого спектру та енергоефективності та, зрештою, більшої ємності для стільникових систем.

Якість. Критерій порівняння полягає в тому, наскільки добре це звучить за ідеальних умов, тобто за так званого чистого аудіо, без помилок передачі та лише з кодуванням.

Поріг імовірності побітової помилки. Вищий поріг помилки призводить до більш надійної структури системи, що призводить до нижчих вимог до співвідношення СШ і збільшення пропускної здатності мережі.

Затримка кодування передаючої системи звуку є фактором, тісно пов'язаним з якісними характеристиками системи.

Таблиця 1.1

Алгоритми кодування

Кодер	Швидкість, кбіт/с	Затримка, мс	Продуктив-ність процесора, MIPS	УЕО
ІКМ з компандуванням	64	0,5	0,2	4,3
ДІКМ	40	1	3	4,1
АДІКМ	32	1	5,6	4,1
Вокодер LD-CELP	16	5	16	3,6
Вокодер GSM RPE	13	15	10	3,8
Вокодер CELP	8	20	15	3,7
Вокодер LPC	2,4	20	10	2,5

1.7 Висновки до розділу 1

Розглянуто особливості обміну інформацією, де основним носійним сигналом є мовні чи голосові сигнали. Показано важливість перетворення таких сигналів для збільшення швидкості обміну даними в цифрових системах зв’язку. При цьому застосовуються спеціальні пристрої – вокодери, які призначені для стиснення ГС, зокрема із врахуванням психоакустичного ефекту маскування.

Проаналізовано особливості голосу для ефективного стиснення таких сигналів в різного типу вокодерах.

Проаналізовано особливості функціонування різного типу вокодерів.

Показано ефективність застосування класу фазових вокодерів, що використовують алгоритм, який може інтерполювати інформацію, в різних областях аудіосигналів, використовуючи інформацію про фазу, отриману з частотного перетворення. Комп’ютерний алгоритм дозволяє модифікувати цифровий звуковий файл у частотній області (зазвичай розширення/стиснення за часом і зсув висоти).

Проаналізувавши недоліки, які притаманні сучасним алгоритмам роботи фазових вокодерів, показано актуальність задачі удосконалення чи розроблення

нових алгоритмів кодування аудіосигналів, зокрема і голосових, для побудови фазових вокодерів.

РОЗДІЛ 2
ОСНОВНА ЧАСТИНА

2.1 Методи синтезу мовних сигналів, як завершальний етап роботи вокодерів

Синтез мовлення — це штучне створення людського мовлення. Комп’ютерна система, яка використовується для цієї мети, називається синтезатором мовлення і може бути реалізована в програмних або апаратних продуктах. Система перетворення тексту в мову (TTS) перетворює звичайний мовний текст у мовлення; інші системи передають символічні лінгвістичні уявлення, такі як фонетична транскрипція, у мову. Зворотний процес — розпізнавання мови.

Синтезоване мовлення можна створити шляхом об’єднання фрагментів записаного мовлення, які зберігаються в базі даних. Системи відрізняються розміром збережених мовних одиниць; система, яка зберігає телефони або дифони, забезпечує найбільший вихідний діапазон, але може не мати чіткості. Для конкретних доменів використання зберігання цілих слів або речень забезпечує високоякісний вихід. Крім того, синтезатор може включати модель голосового тракту та інші характеристики людського голосу для створення повністю «синтетичного» голосового виходу.

Якість синтезатора мовлення оцінюється за його схожістю з людським голосом і здатністю бути чітким для розуміння. Зрозуміла програма перетворення тексту в мовлення дозволяє людям із вадами зору або проблемами читання слухати письмові слова на домашньому комп’ютері. З початку 1990-х років багато комп’ютерних операційних систем включають синтезатори мови.

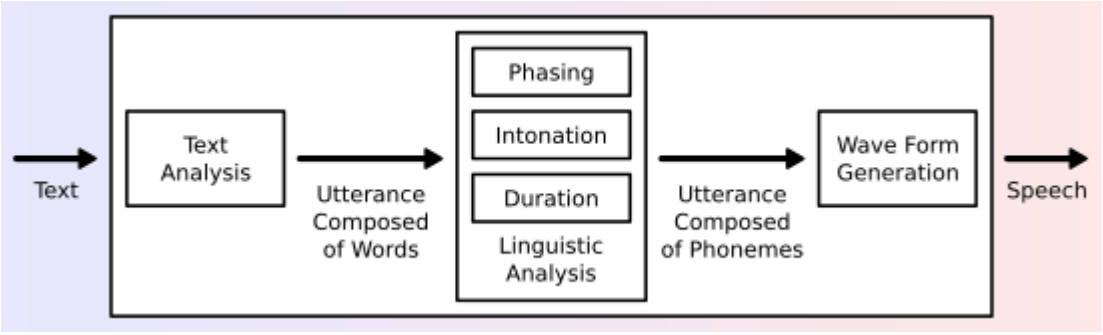

Рис. 2.1. Огляд типової системи TTS

Система перетворення тексту в мовлення (або «мобільний механізм») складається з передньої та задньої частини. Інтерфейс має два основних завдання. По-перше, він перетворює необроблений текст, що містить такі символи, як цифри та абревіатури, на еквівалент виписаних слів. Цей процес часто називають нормалізацією тексту, попередньою обробкою або маркеруванням. Потім інтерфейс призначає фонетичну транскрипцію кожному слову, а також ділить і позначає текст на просодичні одиниці, такі як фрази, речення. Процес присвоєння фонетичної транскрипції словам називається перетворенням тексту у фонеми або перетворенням графеми у фонеми. Фонетична транскрипція та інформація про просодію разом складають символічне лінгвістичне представлення, яке виводиться інтерфейсом. Потім бек-енд, який часто називають синтезатором, перетворює символічне лінгвістичне представлення в звук. У деяких системах ця частина включає в себе обчислення цільової просодії (контур висоти, тривалість фонем), яка потім накладається на вихідну мову.

У 1930-х роках Bell Labs розробила вокодер, який автоматично аналізував мову на основні тони та резонанси. На основі своєї роботи над вокодером Гомер Дадлі розробив клавіатурний синтезатор голосу під назвою The Voder (Демонстратор голосу), який він виставив на Всесвітній виставці в Нью-Йорку 1939 року.

Доктор Франклін С. Купер і його колеги з Haskins Laboratories створили відтворення шаблонів наприкінці 1940-х років і завершили його в 1950 році. Існувало кілька різних версій цього апаратного пристрою; наразі зберігся лише

один. Машина перетворює зображення акустичних моделей мови у формі спектрограми назад у звук. Використовуючи цей пристрій, Елвін Ліберман і його колеги виявили акустичні сигнали для сприйняття фонетичних сегментів (приголосних і голосних).

Рис. 2.2. Корпус комп’ютера та синтезатора мови, який використовував Стівен Гокінг у 1999 році

Лінійне прогностичне кодування (LPC), форма кодування мовлення, почалося завдяки роботі Фумітади Ітакури з Університету Нагоя та Шузо Сайто з Nippon Telegraph and Telephone (NTT) у 1966 році. Подальший розвиток технології LPC зробив Бішну С. Атал. і Манфред Р. Шредер у Bell Labs у 1970-х роках. Пізніше LPC став основою для перших чіпів синтезатора мови, таких як мовні чіпи Texas Instruments LPC, які використовувалися в іграшках Speak & Spell з 1978 року.

У 1975 році Фумітада Ітакура розробив метод лінійних спектральних пар (LSP) для кодування мови з високим ступенем стиснення, перебуваючи в NTT. З 1975 по 1981 рік Ітакура вивчав проблеми аналізу та синтезу мови на основі методу LSP. У 1980 році його команда розробила мікросхему синтезатора мови на основі LSP. LSP є важливою технологією для синтезу та кодування мовлення, і в 1990-х роках вона була прийнята майже всіма міжнародними стандартами кодування мовлення як важливий компонент, що сприяє вдосконаленню цифрового мовного зв’язку через мобільні канали та Інтернет.

У 1975 році була випущена MUSA, яка була системою синтезу мовлення. Вона складався з автономного комп’ютерного обладнання та спеціального програмного забезпечення, яке дозволяло читати італійською мовою. Друга версія, випущена в 1978 році, також могла співати італійською мовою.

Домінуючими системами в 1980-х і 1990-х роках були: система DECtalk, заснована в основному на роботі Денніса Клатта з Массачусетського технологічного інституту, і система Bell Labs.

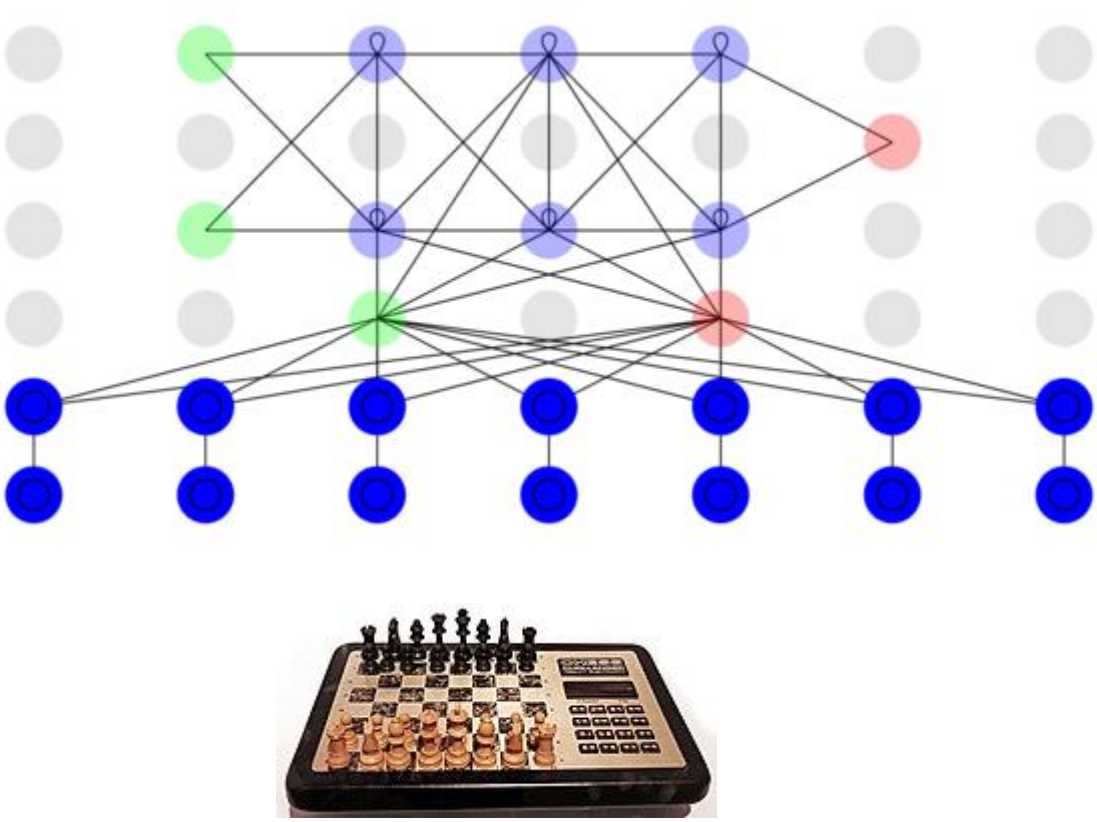

Рис. 2.3. Fidelity Voice Chess Challenger (1979), перший розмовний шаховий комп’ютер

Портативна електроніка з функціями синтезу мови почала з’являтися в 1970-х роках. Одним із перших був портативний калькулятор для сліпих Telesensory Systems Inc. (TSI) Speech+ у 1976 році. Інші пристрої мали в основному освітні цілі, наприклад, іграшка Speak & Spell, вироблена Texas Instruments у 1978 році. У 1979 році компанія Fidelity випустила розмовну версію свого електронного шахового комп’ютера. Першою відеогрою з синтезом

мовлення була аркадна гра Stratovox (відома в Японії як Speak & Rescue) 1980 року від Sun Electronics. Першою персональною комп'ютерною грою з синтезом мовлення була Manbiki Shoujo (Shopolifting Girl), випущена в 1980 році для PET 2001, для якої розробник гри, Хіроші Сузукі, розробив техніку програмування «нульового хреста» для створення синтезованої форми мовлення. Інший ранній приклад, аркадна версія Berzerk, також датується 1980 роком. Компанія Milton Bradley випустила першу багатокористувацьку електронну гру з використанням голосового синтезу, Milton, у тому ж році.

У 1976 році компанія Computalker Consultants випустила свій синтезатор мови CT-1. Розроблений Д. Ллойдом Райсом і Джимом Купером, це був аналоговий синтезатор, створений для роботи з мікрокомп'ютерами, що використовують стандарт шини S-100.

Ранні електронні синтезатори мови звучали як роботизовані і часто були ледь розбірливими. Якість синтезованого мовлення неухильно покращувалася, але станом на 2016 рік вихід із сучасних систем синтезу мовлення залишається чітко відмінним від справжньої людської мови.

Синтезовані голоси зазвичай звучали чоловічими до 1990 року, коли Енн Сірдал з AT&T Bell Laboratories створила жіночий голос.

У 2005 році Курцвейл передбачив, що оскільки співвідношення ціни та ефективності призвело до того, що синтезатори мови стануть дешевшими та доступнішими, більше людей отримають вигоду від використання програм перетворення тексту в мову.

2.2 Синтезаторні технології

Найважливішими якостями системи синтезу мовлення є природність і зрозумілість. Природність описує, наскільки сигнал звучить як людська мова, а розбірливість – це легкість, з якою сигнал розуміється. Ідеальний синтезатор мови є природним і зрозумілим. Системи синтезу мови зазвичай намагаються максимізувати обидві характеристики.

Двома основними технологіями, що генерують сигнали синтетичної мови, є конкатенативний синтез і формантний синтез. Кожна технологія має сильні та слабкі сторони, і передбачуване використання системи синтезу зазвичай визначатиме, який підхід використовується.

Конкатенаційний синтез

Конкатенативний синтез заснований на конкатенації (з'єднанні разом) сегментів записаної мови. Як правило, конкатенативний синтез створює синтезоване мовлення, що звучить найбільш природно. Однак відмінності між природними варіаціями мовлення та характером автоматизованих методів сегментації сигналів іноді призводять до чутних збоїв. Існує три основних підтипи конкатенативного синтезу.

Синтез вибору одиниць

Синтез вибору одиниць використовує великі бази даних записаної мови. Під час створення бази даних кожне записане висловлювання сегментується на деякі окремі: окремі дзвінкі, двоголосні, напівзвукові, склади, морфеми, слова, фрази та речення. Як правило, поділ на сегменти виконується за допомогою спеціально модифікованого розпізнавання мовлення, налаштованого на режим «примусового вирівнювання» з деякою ручною корекцією згодом, використовуючи візуальні представлення, такі як форма хвилі та спектрограма.ьПотім створюється індекс одиниць у базі даних мовлення на основі сегментації та акустичних параметрів, таких як основна частота (висота), тривалість, позиція в складі та сусідні звуки. Під час виконання бажане цільове висловлювання створюється шляхом визначення найкращого ланцюжка одиниць-кандидатів із бази даних (вибір одиниць). Цей процес зазвичай досягається за допомогою спеціально зваженого дерева рішень.

Вибір одиниць забезпечує найбільшу природність, оскільки він застосовує лише невелику кількість даних обробки цифрового сигналу (DSP) до записаної мови. DSP часто робить записане мовлення менш природним, хоча деякі системи використовують невелику кількість обробки сигналу в точці конкатенації для згладжування форми сигналу. Вихідні дані найкращих систем вибору одиниць часто неможливо відрізнити від реальних людських голосів, особливо в

контекстах, для яких налаштована система TTS. Проте максимальна природність зазвичай вимагає, щоб мовні бази даних із вибором одиниць були дуже великими, у деяких системах сягаючи гігабайтів записаних даних, що представляє десятки годин мовлення. Крім того, відомо, що алгоритми вибору одиниць вибирають сегменти з місця, що призводить до не ідеального синтезу (наприклад, другорядні слова стають незрозумілими), навіть якщо в базі даних існує кращий вибір. Нещодавно дослідники запропонували різні автоматизовані методи виявлення неприродних сегментів у системах синтезу мови з вибором одиниць.

Дифонний синтез

Синтез дифонів використовує мінімальну базу даних мовлення, що містить усі дифони (переходи звуку в звук), що відбуваються в мові. Кількість дифонів залежить від фонотактики мови: наприклад, в іспанській мові близько 800 дифонів, а в німецькій – близько 2500. При синтезі дифонів лише один приклад кожного дифону міститься в базі даних мовлення. Під час виконання цільова просодія речення накладається на ці мінімальні одиниці за допомогою цифрових методів обробки сигналів, таких як лінійне прогнозне кодування, PSOLA або MBROLA. чи новіші методи, такі як модифікація висоти у вихідній області за допомогою дискретного косинусного перетворення. Дифонний синтез страждає від звукових збоїв конкатенативного синтезу та роботизованої природи формантного синтезу, і має лише деякі переваги обох підходів, окрім невеликого розміру. Таким чином, його використання в комерційних програмах знижується, хоча він продовжує використовуватися в дослідженнях, оскільки існує низка вільнодоступних реалізацій програмного забезпечення.

Доменно-орієнтований синтез

Доменно-орієнтований синтез об’єднує попередньо записані слова та фрази для створення повних висловлювань. Він використовується в додатках, де різноманітність текстів, які виводить система, обмежена певним доменом, як-от повідомлення про розклад руху транспорту чи звіти про погоду. Цю технологію дуже просто реалізувати, і вона вже давно використовується в комерційних цілях у таких пристроях, як розмовні годинники та калькулятори. Рівень природності

цих систем може бути дуже високим, оскільки різноманітність типів речень обмежена, і вони точно відповідають просодії та інтонації оригінальних записів.

Оскільки ці системи обмежені словами та фразами у своїх базах даних, вони не є універсальними та можуть синтезувати лише комбінації слів та фраз, на які вони були попередньо запрограмовані. Однак змішування слів у природній розмовній мові все ще може спричинити проблеми, якщо не взяти до уваги численні варіації.

Формантний синтез

Формантний синтез не використовує зразки людської мови під час виконання. Натомість синтезований мовний вихідний сигнал створюється за допомогою адитивного синтезу та акустичної моделі (синтез фізичного моделювання). Такі параметри, як основна частота, голос і рівні шуму, змінюються з часом для створення хвилі штучної мови. Цей метод іноді називають синтезом на основі правил; однак багато конкатенативних систем також мають компоненти, засновані на правилах. Багато систем, заснованих на технології синтезу формант, генерують штучну мову, що звучить роботом, яку ніколи не сплутаєш з людською. Однак максимальна природність не завжди є метою системи синтезу мовлення, і системи формантного синтезу мають переваги перед конкатенативними системами. Синтезоване формантне мовлення може бути надійно зрозумілим навіть на дуже високих швидкостях, уникаючи акустичних збоїв, які зазвичай викликають конкатенаційні системи. Високошвидкісне синтезоване мовлення використовується людьми з вадами зору для швидкої навігації по комп’ютерах за допомогою програми зчитування з екрана. Формантні синтезатори зазвичай є меншими програмами, ніж конкатенативні системи, оскільки вони не мають бази даних зразків мовлення. Таким чином придатні до функціонування у вбудованих системах, де потужність пам’яті та мікропроцесора особливо обмежена. Оскільки системи, засновані на формантах, мають повний контроль над усіма аспектами вихідного мовлення, можна виводити широкий спектр просодії та інтонацій, передаючи не лише запитання та твердження, але різноманітні емоції та відтінки голосу.

Артикуляційний синтез складається з обчислювальних прийомів синтезу мови на основі моделей голосового апарату людини та процесів артикуляції, що відбуваються в ньому. Перший артикуляційний синтезатор, який регулярно використовувався для лабораторних експериментів, був розроблений у Haskins Laboratories у середині 1970-х років Філіпом Рубіном, Томом Бером і Полом Мермельстайном. Цей синтезатор, відомий як ASY, базувався на моделях голосового тракту, розроблених у Bell Laboratories у 1960-х і 1970-х роках Полом Мермельстайном, Сесілом Кокером та їх колегами.

До недавнього часу моделі артикуляційного синтезу не були включені в комерційні системи синтезу мовлення. Помітним винятком є система на основі NeXT, спочатку розроблена та продана компанією Trillium Sound Research, дочірньою компанією Університету Калгарі, де було проведено більшість оригінальних досліджень. Після припинення різноманітних втілень NeXT (започаткованих Стівом Джобсом наприкінці 1980-х і об’єднаних з Apple Computer у 1997 році) програмне забезпечення Trillium було опубліковано під ліцензією GNU General Public License, а робота продовжувалася як gnuspeech. Система, вперше представлена на ринку в 1994 році, забезпечує повне перетворення тексту в мову на основі артикуляції за допомогою хвилеводу або аналогової лінії передачі ротових і носових шляхів людини, керованих «моделлю характерної області».

Більш сучасні синтезатори, розроблені Хорхе К. Лусеро та його колегами, включають моделі біомеханіки голосових складок, аеродинаміки голосової складки та поширення акустичних хвиль у бронхах, трахеї, носовій та ротовій порожнинах і, таким чином, становлять повну систему симуляції мовлення на основі фізики.

Синтез на основі HMM

Синтез на основі HMM — це метод синтезу, заснований на прихованих моделях Маркова, також званий статистичним параметричним синтезом. У цій системі частотний спектр (голосовий тракт), основна частота (джерело голосу) і тривалість (просодія) мови моделюються одночасно HMM. Сигнали мовлення генеруються з самих HMM на основі критерію максимальної правдоподібності.

Синтез мовлення глибокого навчання використовує глибокі нейронні мережі (DNN) для створення штучного мовлення з тексту (синтез мовлення) або спектру (вокодер). Глибокі нейронні мережі навчаються з використанням великої кількості записаного мовлення та, у випадку системи перетворення тексту в мовлення, пов’язаних міток та/або введеного тексту.

15.ai використовує модель із кількома динаміками — сотні голосів навчаються одночасно, а не послідовно, зменшуючи необхідний час навчання та дозволяючи моделі вивчати та узагальнювати спільний емоційний контекст, навіть для голосів, які не піддаються такому емоційному контексту. Модель глибокого навчання, яку використовує програма, є недетермінованою: кожного разу, коли мова генерується з того самого рядка тексту, інтонація мови буде дещо іншою. Програма також підтримує вручну змінювати емоцію створеного рядка за допомогою емоційних контекстуалізаторів (термін, введений у рамках цього проекту), речення чи фрази, які передають емоцію, яка служить орієнтиром для моделі під час генерації.

ElevenLabs в першу чергу відома своїм програмним забезпеченням Speech Synthesis, заснованим на браузері та за допомогою штучного інтелекту, який може створювати реалістичне мовлення, синтезуючи вокальні емоції та інтонацію. Компанія стверджує, що її програмне забезпечення створено для налаштування інтонації та темпу подачі на основі контексту використовуваної мови. Він використовує розширені алгоритми для аналізу контекстуальних аспектів тексту з метою виявлення таких емоцій, як гнів, смуток, щастя чи тривога, що дає змогу системі зрозуміти настрої користувача, що призводить до більш реалістичного та схожого на людину виразу. Інші функції включають багатомовну генерацію мовлення та створення довгоформатного контенту за допомогою голосів з урахуванням контексту.

Синтезатори мови на основі DNN наближаються до природності людського голосу. Прикладами недоліків методу є низька надійність, коли даних недостатньо, відсутність керованості та низька продуктивність авторегресійних моделей.

Технологія audio deepfake, яку також називають клонуванням голосу або deepfake audio, — це застосування штучного інтелекту, призначене для створення мови, яка переконливо імітує конкретних людей, часто синтезуючи фрази чи речення, які вони ніколи не говорили. Спочатку розроблений з наміром покращити різні аспекти людського життя, він має практичні застосування, такі як створення аудіокниг і допомога особам, які втратили голос через захворювання. Крім того, він має комерційне використання, включаючи створення персоналізованих цифрових помічників, систем перетворення тексту в мовлення з природним звучанням і передових служб перекладу мовлення.

2.3 Перетворення тексту на фонему

Системи синтезу мовлення використовують два основні підходи для визначення вимови слова на основі його написання, процес, який часто називають перетворенням тексту у фонему або графему у фонему (фонема — це термін, який лінгвісти використовують для опису характерних звуків у мові). Найпростішим підходом до перетворення тексту у фонему є підхід на основі словника, коли програма зберігає великий словник, що містить усі слова мови та їхню правильну вимову. Визначення правильної вимови кожного слова полягає в пошуку кожного слова в словнику та заміні написання на вимову, вказану в словнику. Інший підхід заснований на правилах, коли правила вимови застосовуються до слів, щоб визначити їхню вимову на основі їх написання. Це схоже на підхід до навчання читання «озвучуванням», або синтетичною акустикою.

Кожен підхід має переваги та недоліки. Підхід, заснований на словнику, є швидким і точним, але повністю зазнає невдачі, якщо дається слово, якого немає в його словнику. Зі збільшенням розміру словника зростають і вимоги до пам’яті системи синтезу. З іншого боку, підхід, заснований на правилах, працює з будь-яким введенням, але складність правил суттєво зростає, оскільки система враховує неправильне написання або вимову. У результаті майже всі системи синтезу мовлення використовують комбінацію цих підходів.

Послідовна оцінка систем синтезу мовлення може бути важкою через відсутність загальновизнаних об’єктивних критеріїв оцінки. Різні організації часто використовують різні мовні дані. Якість систем синтезу мови також залежить від якості техніки, яка може включати аналоговий або цифровий запис, і засобів, що використовуються для відтворення мови. Таким чином, оцінка систем синтезу мовлення часто була складною через відмінності між технікою генерації та засобами відтворення.

Синтез мовлення (TTS) стосується здатності комп’ютерів читати текст вголос. Механізм TTS перетворює письмовий текст на фонематичне представлення, а потім перетворює фонематичне представлення на сигнали, які можна вивести як звук. Механізми TTS з різними мовами, діалектами та спеціалізованими словниками вже є доступні.

2.4 Розтягнення часу аудіо та масштабування висоти

Розтягнення часу — це процес зміни швидкості або тривалості звукового сигналу без впливу на його висоту. Масштабування висоти – це навпаки: процес зміни висоти без впливу на швидкість. Pitch shift — це масштабування висоти, реалізоване в блоці ефектів і призначене для живого виконання. Керування висотою звуку — це простіший процес, який одночасно впливає на висоту тону та швидкість, уповільнюючи або прискорюючи запис.

Ці процеси часто використовуються для узгодження висоти та темпу двох попередньо записаних кліпів для мікшування, коли кліпи не можна відтворити повторно або пересемплювати. Розтягнення часу часто використовується для коригування радіореклами та аудіо телевізійної реклами, щоб точно відповідати доступним тривалостям. Його можна використовувати для узгодження довшого матеріалу з визначеним часовим проміжком, наприклад, для 1-годинної трансляції.

Повторна вибірка

Найпростіший спосіб змінити тривалість або висоту звуку — змінити швидкість відтворення. Для цифрового аудіозапису це можна зробити за

допомогою перетворення частоти дискретизації. Під час використання цього методу частоти в записі завжди масштабуються в тому самому співвідношенні, що й швидкість, транспонуючи її сприйнятий тон угору або вниз у процесі. Уповільнення запису для збільшення тривалості також знижує висоту звуку, тоді як пришвидшення на менший час відповідно підвищує висоту звуку. Під час повторної дискретизації аудіо до значно нижчого тону може бути бажаним, щоб вихідний аудіо мав вищу частоту дискретизації, оскільки уповільнення частоти відтворення призведе до відтворення аудіосигналу з нижчою роздільною здатністю, а отже, зменшить сприйману чіткість звуку. Навпаки, під час передискретизації аудіо до помітно вищого тону, можливо, буде краще включити інтерполяційний фільтр, оскільки частоти, які перевищують частоту Найквіста, зазвичай створюватимуть небажані спотворення звуку, явище, яке також називають псевдонімом.

2.5 Техніка фазового кодування

Один із способів збільшити довжину сигналу, не впливаючи на висоту тону, — побудувати фазовий вокодер за Фланаганом, Голденом.

Основні кроки:

- обчислити миттєве співвідношення частота/амплітуда сигналу за допомогою STFT, яке є дискретним перетворенням Фур’є короткого блоку вибірок, що перекривається та має плаваюче вікно;

- застосувати деяку обробку до величин і фаз перетворення Фур’є (наприклад, повторна дискретизація блоків ШПФ); і

- виконати зворотне STFT, виконавши зворотне перетворення Фур’є для кожного фрагменту та додавши отримані фрагменти сигналу, що також називається перекриттям і додаванням (OLA).

Фазовий вокодер добре обробляє компоненти синусоїди, але ранні реалізації ввели значне розмиття на перехідних («биттях») формах сигналу при всіх нецілочисельних швидкостях стиснення/розширення, що робить результати фазовими та дифузними. Останні вдосконалення забезпечують кращу якість

результатів при всіх ступенях стиснення/розширення, але залишковий ефект розмазування все ще залишається.

Техніка фазового вокодера також може бути використана для виконання зсуву висоти звуку, маніпуляції тембром, гармонізації та інших незвичайних модифікацій, які можна змінювати залежно від часу.

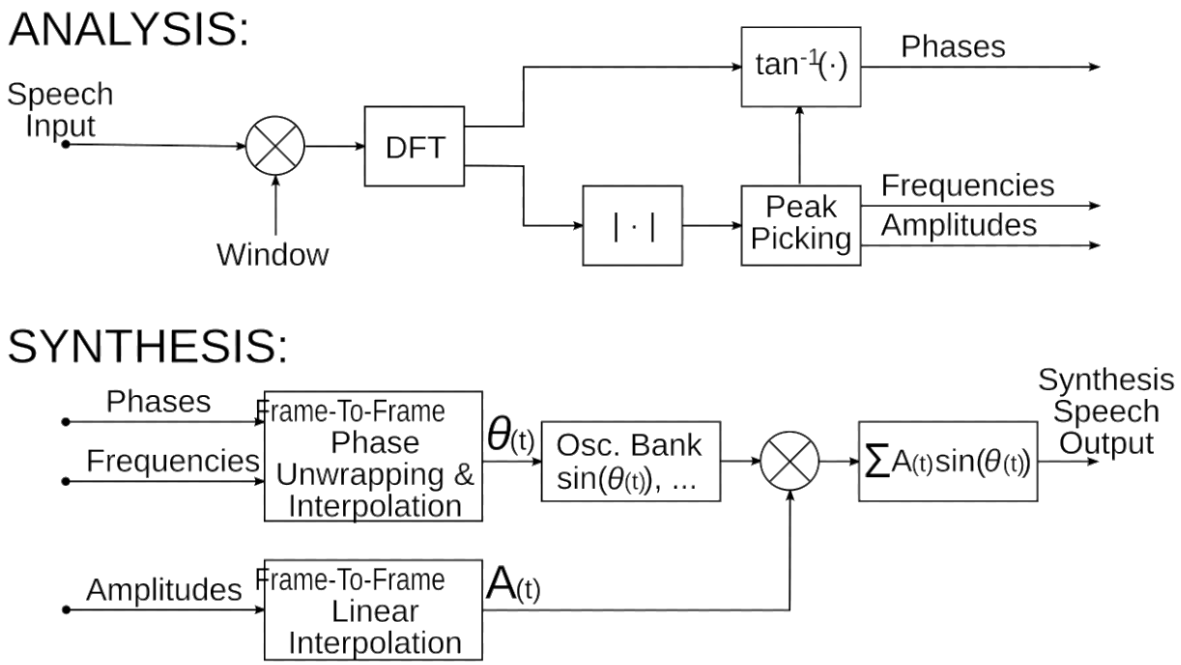

Рис. 2.4. Синусоїдальна система аналізу/синтезу

Синусоїдальне спектральне моделювання

Інший метод розтягування часу базується на спектральній моделі сигналу. У цьому методі піки ідентифікуються в кадрах за допомогою STFT сигналу, а синусоїдальні «доріжки» створюються шляхом з’єднання піків у сусідніх кадрах. Потім треки повторно синтезуються в новому часовому масштабі. Цей метод може дати хороші результати як для поліфонічного, так і для ударного матеріалу, особливо коли сигнал розділений на піддіапазони. Однак цей метод вимагає більше обчислень, ніж інші методи.

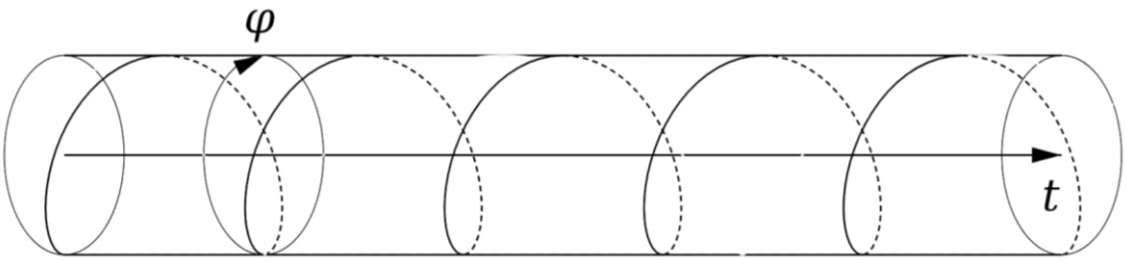

Рис. 2.5. Моделювання одноголосного звуку як спостереження вздовж спіралі функції з циліндричною областю

Часова область

Рабінер і Шафер у 1978 році запропонували альтернативне рішення, яке працює в часовій області: спроба знайти період (або, еквівалентно, основну частоту) заданої ділянки хвилі, використовуючи певний алгоритм виявлення висоти (зазвичай пік автокореляції сигналу, або іноді кепстральна обробка), і переводити один період в інший.

Це називається гармонічним масштабуванням у часовій області або методом синхронізованого перекривання-додавання (SOLA) і працює дещо швидше, ніж фазовий вокодер на повільніших машинах, але не працює, коли автокореляція неправильно оцінює період сигналу зі складними гармоніками.

Adobe Audition (раніше Cool Edit Pro) вирішує це, шукаючи період, найближчий до центрального періоду, який визначає користувач, який має бути цілим числом, кратним темпу, і між 30 Гц і найнижчою частотою низьких частот.

Це значно обмеженіше за сферою застосування, ніж обробка на основі фазового вокодера, але її можна зробити набагато менш інтенсивною для процесора для програм реального часу. Він забезпечує найбільш узгоджені результати для однотональних звуків, як-от записи голосу чи монофонічних музичних інструментів.

Висококласні комерційні пакети обробки аудіо або поєднують обидві методи (наприклад, шляхом поділу сигналу на синусоїдну та перехідну форми хвилі), або використовують інші методи, засновані на вейвлет-перетворенні, або обробці штучної нейронної мережі, проводячи найвищу якіст розтягування часу.

Фреймовий підхід

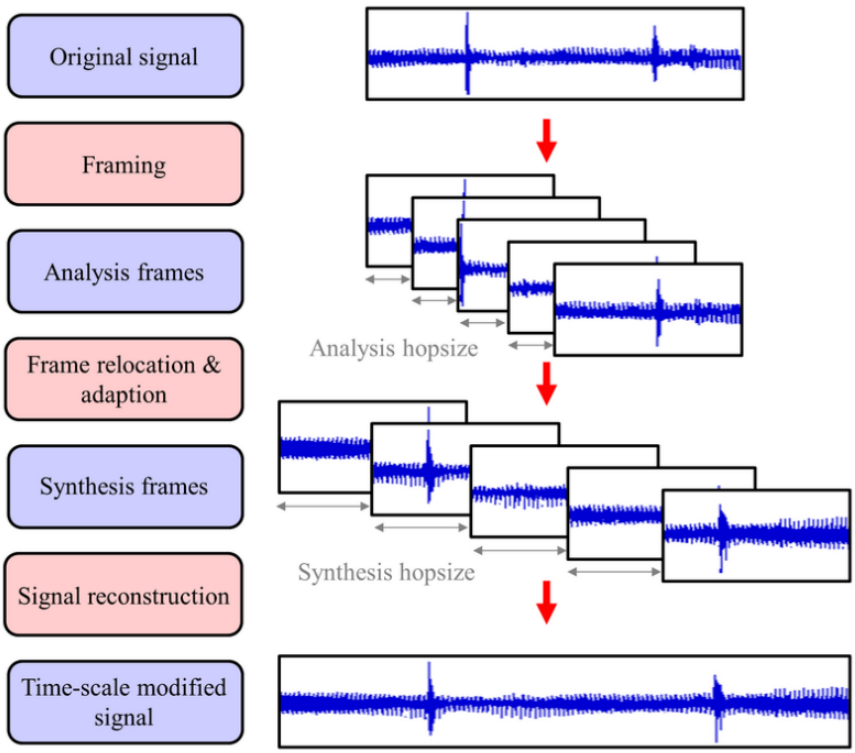

Рис. 2.6. Фреймовий підхід багатьох процедур TSM

Щоб зберегти висоту звукового сигналу під час розтягування або стиснення його тривалості, багато процедур модифікації шкали часу (TSM) використовують підхід, заснований на кадрах. Враховуючи оригінальний аудіосигнал у дискретному часі, першим кроком цієї стратегії є розділення сигналу на короткі кадри аналізу фіксованої довжини. Кадри аналізу розділені фіксованою кількістю зразків, яка називається стрибкоподібним розміром аналізу Ha∈N. Щоб досягти фактичної модифікації масштабу часу, кадри аналізу потім тимчасово переміщуються, щоб отримати стрибкоподібний розмір синтезу Hs∈N. Це переміщення кадру призводить до модифікації тривалості сигналу на коефіцієнт розтягування α=Hs/Ha. Однак просте накладання немодифікованих кадрів аналізу зазвичай призводить до небажаних артефактів, таких як фазові розриви або коливання амплітуди. Щоб запобігти подібним артефактам, кадри аналізу адаптовані для формування кадрів синтезу перед реконструкцією вихідного сигналу, модифікованого за шкалою часу.

Стратегія отримання кадрів синтезу з кадрів аналізу є ключовою відмінністю між різними процедурами TSM.

Швидкість слуху та швидкість розмови

Для конкретного випадку мовлення розтягнення часу можна виконати за допомогою PSOLA.

Стиснене мовлення — це представлення вербального тексту в стисненому часі. Хоча можна очікувати, що прискорення зменшить розуміння, Герб Фрідман каже, що «експерименти показали, що мозок працює найефективніше, якщо швидкість передачі інформації через вуха — через мову — є «середньою» швидкістю читання, яка становить приблизно 200–300 сл./хв., але середня швидкість мовлення знаходиться в районі 110–170 сл./хв.».

Слухання стиснутого мовлення розглядається як еквівалент швидкісного читання.

Ці методи також можна використовувати для транспонування аудіосемплів, утримуючи постійну швидкість або тривалість. Цього можна отримати розтягуванням часу, а потім повторною вибіркою до вихідної довжини. Як альтернатива, частота синусоїд у синусоїдальній моделі може бути змінена безпосередньо, а сигнал реконструйований у відповідному масштабі часу.

Транспонування можна назвати масштабуванням частоти або зсувом висоти, залежно від ракурсу.

Наприклад, можна збільшити висоту кожної ноти на ідеальну квінту, зберігаючи темп незмінним. Можна розглядати цю транспозицію як «зміщення висоти», «зсув» кожної ноти на 7 клавіш угору на фортепіанній клавіатурі, або додавання фіксованої величини на шкалі Мела, або додавання фіксованої величини в лінійному просторі висоти. Це ж транспонування можна розглядати як «масштабування частоти».

Музичне транспонування зберігає співвідношення гармонійних частот, які визначають тембр звуку, на відміну від частотного зсуву, який виконується за допомогою амплітудної модуляції, яка додає фіксований частотний зсув до частоти кожної ноти. Теоретично можна виконати буквальне масштабування висоти, за якого масштабується розташування музичного простору висоти (вища нота буде зміщена на більший інтервал у лінійному просторі висоти, ніж нижча нота), але це дуже незвично.

Обробка в часовій області тут працює набагато краще, оскільки змазування менш помітне, але масштабування вокальних семплів спотворює спотворені форманти, що може бути бажаним чи небажаним. Процес, який зберігає форманти та характер голосу, передбачає аналіз сигналу канальним вокодером або вокодером LPC плюс будь-який із кількох алгоритмів виявлення висоти тону, а потім його повторний синтез на іншій основній частоті.

2.6 Принципи функціонування фазових вокодерів

Фазовий вокодер — це тип алгоритму, що призначений для вокодера, який може інтерполювати інформацію, наявну в частотній і часовій областях аудіосигналів, використовуючи інформацію про фазу, отриману з частотного перетворення. Комп’ютерний алгоритм дозволяє модифікувати цифровий звуковий файл у частотній області (зазвичай розширення/стиснення за часом і зсув висоти).

В основі фазового вокодера лежить короткочасне перетворення Фур’є (STFT), яке зазвичай кодується за допомогою швидких перетворень Фур’є. STFT перетворює представлення звуку в часовій області в частотно-часове представлення, дозволяючи модифікувати амплітуди або фази певних частотних компонентів звуку перед повторним синтезом представлення частотно-часової області в часову область за допомогою зворотного STFT. Часова еволюція повторно синтезованого звуку може бути змінена за допомогою зміни часової позиції кадрів STFT перед операцією повторного синтезу, що дозволяє модифікувати часову шкалу вихідного звукового файлу.

Основна проблема, яку необхідно вирішити для всіх випадків маніпулювання STFT, визначає, що окремі компоненти сигналу (синусоїди, імпульси) будуть розповсюджені на численні кадри та численні місця розташування частот (бінів) STFT. Тому, аналіз STFT виконується з використанням вікон аналізу, що перекриваються. Вікно призводить до спектрального витоку, так що інформація окремих синусоїдальних компонентів розподіляється по сусідніх вікнах STFT. Щоб уникнути граничних ефектів

звуження вікон аналізу, вікна аналізу STFT перекриваються в часі. Це часове накладання призводить до того, що результати аналізу суміжних вікон STFT сильно корельовані (синусоїда, присутня у кадрі аналізу в момент часу «t», також буде присутня в наступних кадрах). Проблема перетворення сигналу за допомогою фазового вокодера пов’язана з проблемою того, що всі модифікації, які виконуються в поданні STFT, повинні зберігати відповідну кореляцію між сусідніми частотними елементами (вертикальна когерентність) і часовими межами (горизонтальна когерентність). За винятком надзвичайно простих синтетичних звуків, ці відповідні кореляції можуть бути збережені лише приблизно, і з моменту винаходу фазового вокодера дослідження в основному були спрямовані на пошук алгоритмів, які зберегли б вертикальну та горизонтальну узгодженість подання STFT після модифікації. Проблема фазової когерентності досліджувалася досить довго, перш ніж з'явилися відповідні рішення.

Фазовий вокодер був представлений у 1966 році Фланаганом як алгоритм, який зберігає горизонтальну когерентність між фазами бінів, які представляють синусоїдні компоненти. Цей вихідний фазовий вокодер не враховував вертикальну когерентність між сусідніми межами частоти, і тому розтягування часу за допомогою цієї системи створювало звукові сигнали, які були нечіткими.

Оптимальна реконструкція звукового сигналу від STFT після модифікації амплітуди була запропонована Гріффіном і Лімом у 1984 році. Цей алгоритм не розглядає проблему створення когерентного STFT, але він дозволяє знайти звуковий сигнал, який має STFT, максимально близький до модифікованого STFT, навіть якщо модифікований STFT не є когерентним (не представляє жодного сигналу) .

Проблема вертикальної когерентності залишалася основною проблемою для якості операцій масштабування часу до 1999 року, коли Ларош і Долсон запропонували засіб для збереження узгодженості фаз у спектральних межах. Пропозицію Лароша і Долсона слід розглядати як поворотний момент в історії фазового вокодера. Було показано, що за допомогою забезпечення узгодженості вертикальної фази можна отримати дуже якісні перетворення масштабування

часу. Але алгоритм, запропонований Ларошем, не дозволяв зберегти вертикальну фазову когерентність. Рішення цієї проблеми було запропоновано Робелем.

Основне призначення фазового вокодера полягає у розділенні часових і частотних даних. Тому фазові вокодери часто використовуються в музичних програмах.

Першою такою програмою стала програма аналізу звуків музичних інструментів для визначення часових коливань амплітуди та частоти компонентів. З тих пір фазові вокодери знайшли багато різних використань. При цьому виділяються два: масштабування часу та перенесення (транспонування) частоти.

Нехай вокодер виконаний у формі фільтра, сигнал легко розтягнути за часом провівши процедуру інтерполяції сигналу, поданого на вхід генератора / синтезатора вокодера.

Інформація про часові характеристики знаходиься в модульованому сигналі як варіації амплітуди та фази, тоді, як дані про частотні характеристики знаходяться в модульованому сигналі, створюваному генератором (рис.2.7).

Виникнення при цьому «додаткових» вимірювань модульованого сигналу означає, що модульовані частоти є незмінними, але розтягуються в часі.

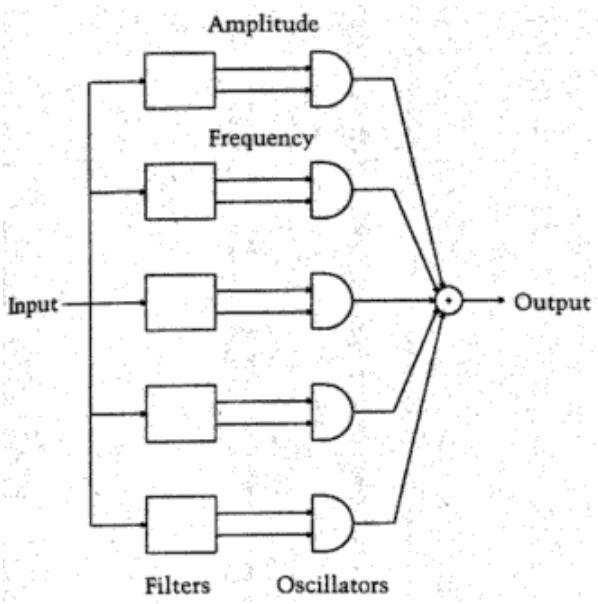

Рис.2.7. Представлення фазового вокодера у формі групи фільтрів

Якщо такий вокодер реалізовано з короткочасним ПФ, часове викривлення сигналу виглядає інакше. Тут етап синтезу грунтується на зворотньому ПФ.

Якщо після цього ми застосовуємо інтерполяцію, тривалість сигналу значно виросте. У той же час, однак, зменшуються частоти гармонік, що утворюють сигнал.

Таким чином, питання удосконалення чи розроблення нових алгоритмів кодування аудіосигналів, зокрема і голосових, для побудови фазових вокодерів, залишається актуальним.

2.7 Висновки до розділу 2

В розділі розглянуто відомі базові дослідження, які проводились для розроблення ефективних алгоритмів кодування аудіосигналів в фазових вокодерах, зокрема розтягування та зтиснення. Показано загальні вимоги та необхідні етапи обробки ГС в фазових вокодерах.

РОЗДІЛ 3
НАУКОВО-ДОСЛІДНА ЧАСТИНА

3.1 Відбір ГС

Для відбору ГС застосовано комп’ютерну стандартні гарнітуру та середовище Adobe Audition (рис. 3.1).

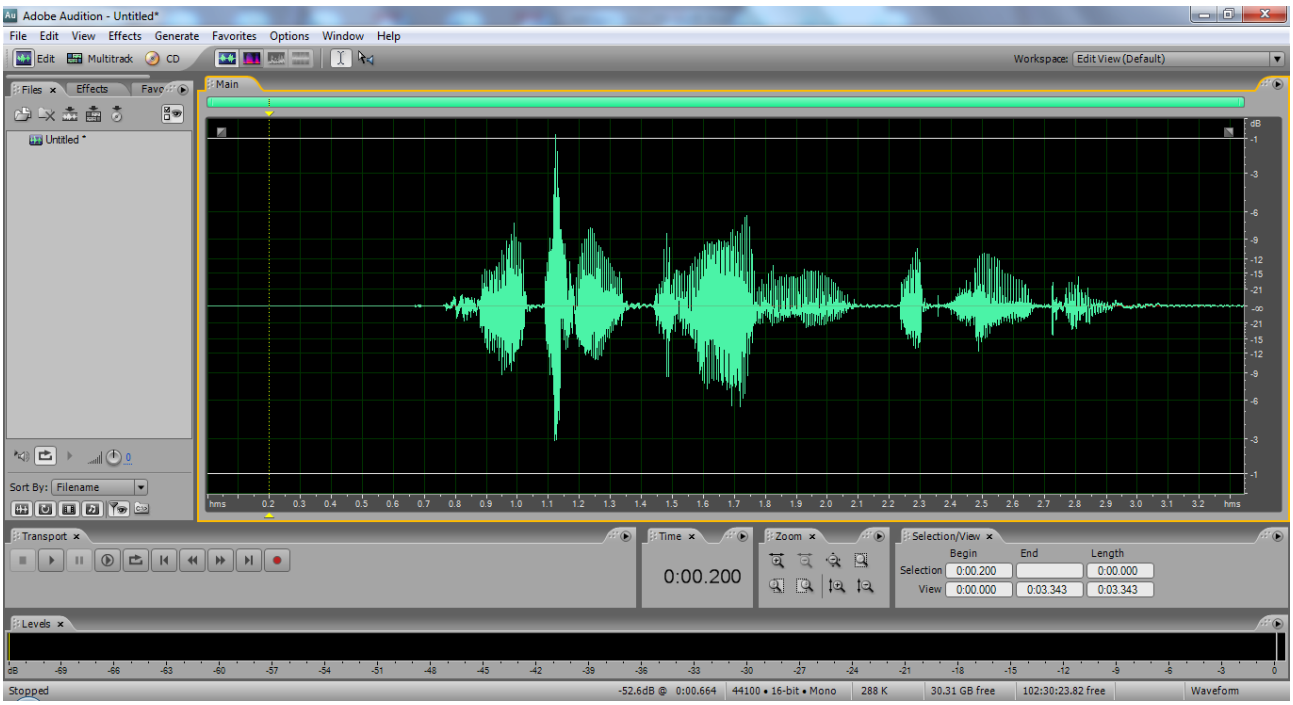

Рис. 3.1. Вікно Adobe Audition

Дальше реєстрований сигнал зберігався та імпортувався в Matlab (рис. 3.2).

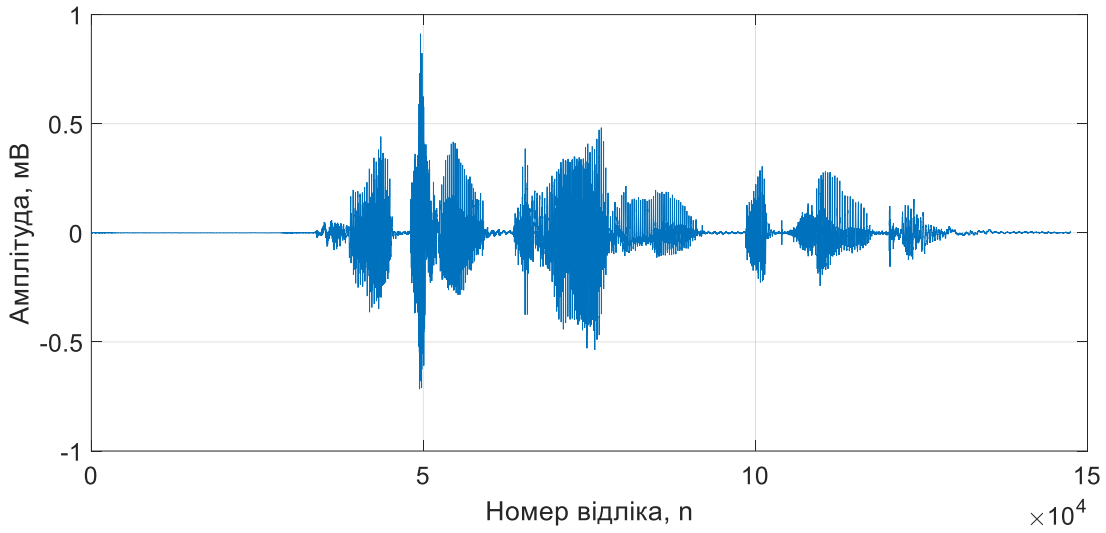

Рис. 3.2. Реєстрограма ГС

На наступному етапі було виконано процес стиснення ГС (рис. 3.2) та наступне його розтягування і розглянуто похибку, яка характеризуватиме сам процес роботи такого вокодера.

3.2 Стиснення ГС

Використовуючи розроблену програму, проведено стиснення ГС вдвічі. Рис. 3.3 відображає результат розтягування записаного попередньо ГС.

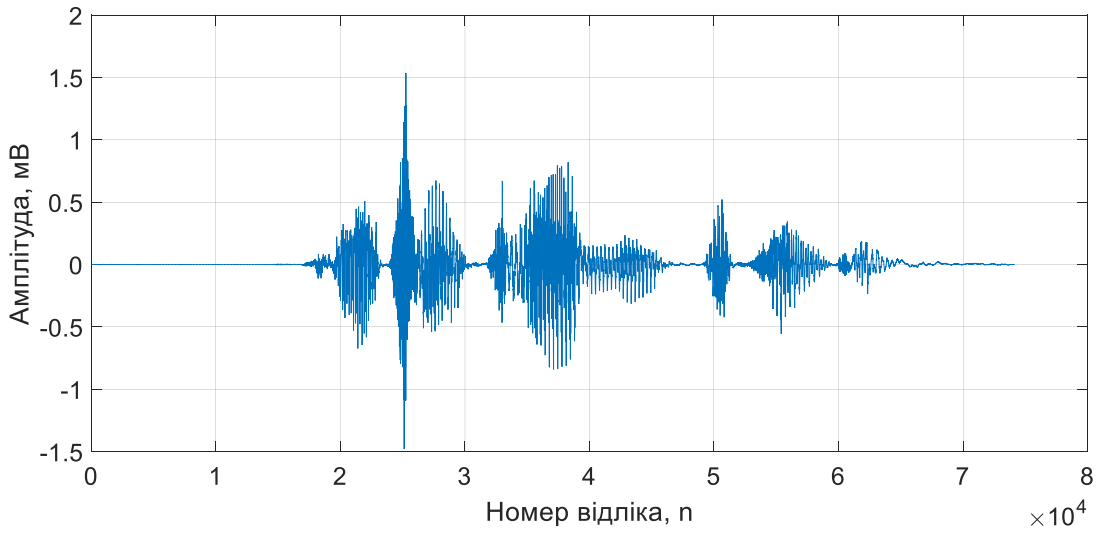

Рис. 3.3. Відображення стиснутого вдвічі ГС

Дальше виконано сповільнення ГС (сигнал показаний на рис. 3.4).

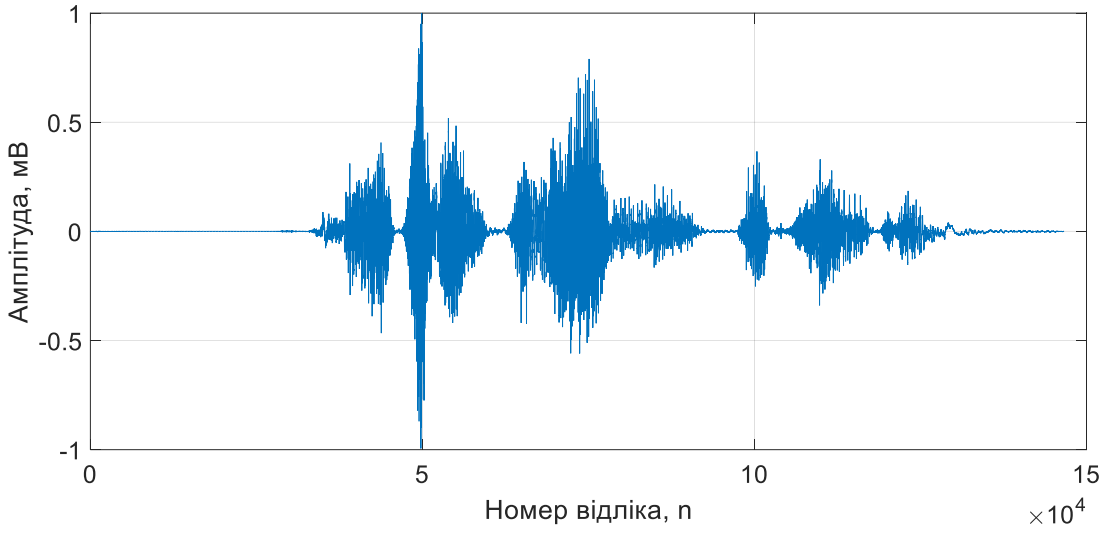

Рис. 3.4. Реєстрограма відновленого ГС

Аналізуючи наведені вище рисунки помічено, що коли сигнал прискорюється, гучність (кількість відліків) зменшується в два рази.

Якщо обсяг вихідного ГС становить 300 Кб, то обсяг прискореного сигналу займає 150 Кб. Тому ГС був стиснутий.

Під час прослуховування ми виявили, що звукова чіткість сигналу залишалася практично однаковою для початкового і відновленого сигналів.

Для визначення похибки реконструкції проаналізовано спектри амплітуд початкового, прискореного та реконструйованого сигналів.

3.3 Оцінювання амплітудних спектрів ГС

Рис. 3.5 відображає спектр амплітуд початкового ГС.

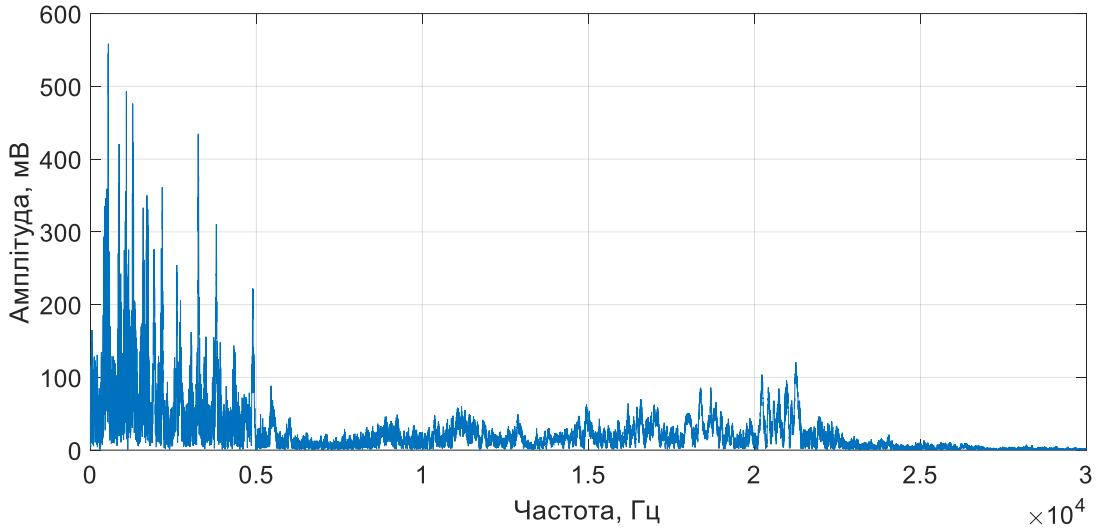

Рис. 3.5. Спектр амплітуд початкового ГС

Рис. 3.6 відображає спектр амплітуд стиснутого ГС, а рис. 3.7 – спектр амплітуд реконструйованого ГС.

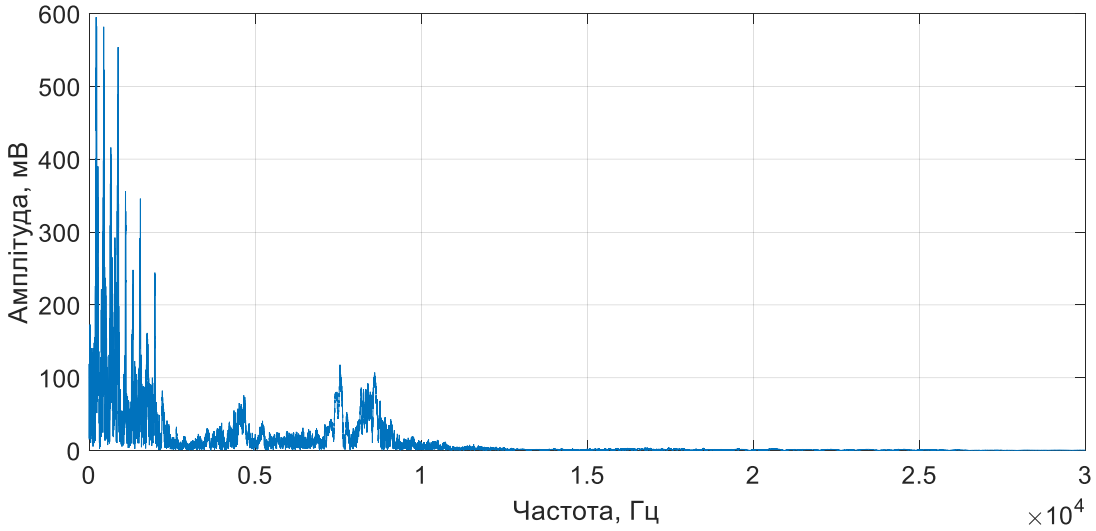

Рис. 3.6. Спектр амплітуд стиснутого ГС

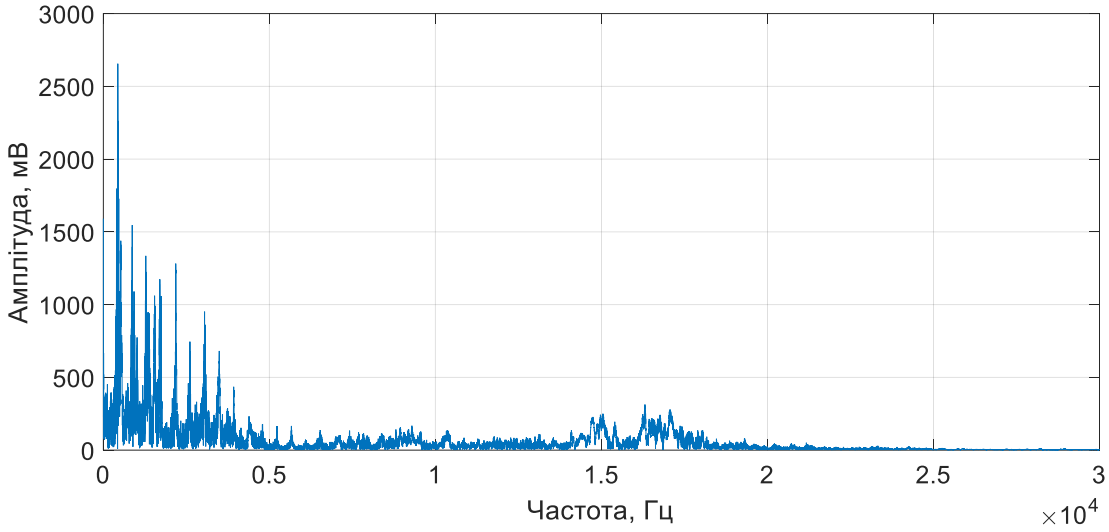

Рис. 3.7. Спектр амплітуд реконструйованого ГС

Як і було прогнозовано, спектр стиснутого ГС звузився вдвічі.

3.4 Імітація ефекту «гармонізації»

Цей цікавий ефект проявляється в тому, що певний текст промовляється ніби двома дикторами з різними голосами.

На рис. 3.8 наведено вигляд отриманої після змін реєстрограми ГС, а на рис. 3.9 – оцінки амплітудного спектру зміненого сигналу.

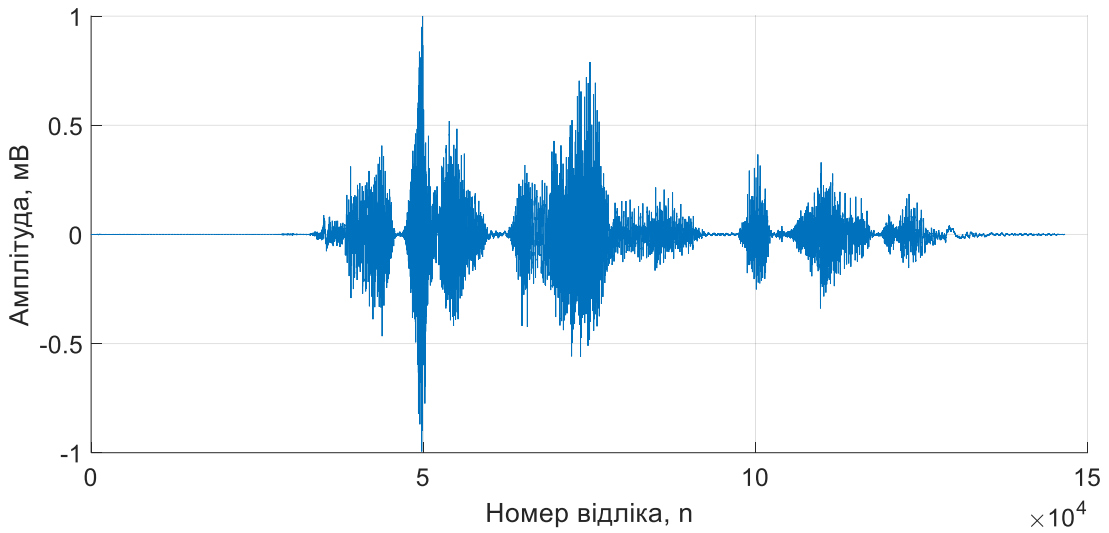

Рис. 3.8. Вигляд отриманої після змін реєстрограми ГС

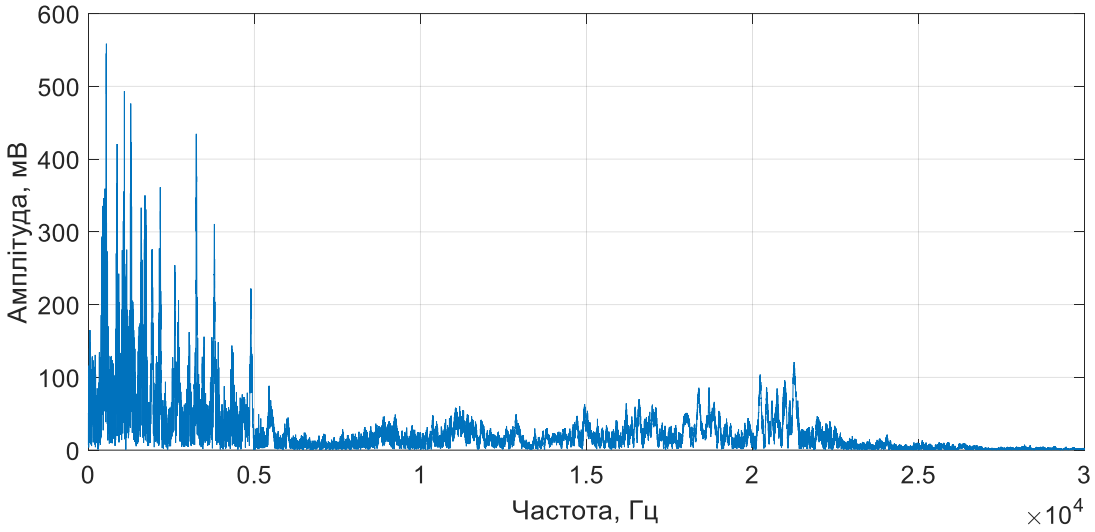

Рис. 3.9. Оцінки амплітудного спектру зміненого сигналу

Порівнюючи візуально вигляд оцінок вихідного та сигналу після гармонізації можна помітити, що вони відрізняються незначно, зате відмінності в звучанні є суттєві. Це дає можливість шифрувати ГС на приклад в системах аутентифікації користувачів. Підробити кодований сигнал майже неможливо, тому що для цього, крім оригінального ГС, знадобиться набір параметрів роботи алгоритму сповільнення/пришвидшення та зміни висоти тону.

3.5 Висновки до розділу 3

Проведено реалізацію основних етапів обробки ГС для фазових фокодерів в середовищі Matlab. При цьому виконано аудіозапис з допомогою програми Adobe Audition 3.0 та стандартного комп’ютерного мікрофона і наступне опрацювання його в середовищі Matlab.

Попередньо було проведено низькочастотну фільтрацію цього сигналу. Як алгоритм опрацювання використано підходи, описані в працях Ларош і Долсона, що передбачає часове стиснення/розтягування аудіосигналу, а також зміну висоти голосу диктора в аудіофайлі.

Проведено моделювання процесу стиснення ГС та наступного його розтягування і оцінено похибку, яка характеризує процес кодування/декодування сигналу в фазовому вокодері.

Під час прослуховування виявлено, що звукова чіткість сигналу залишалася практично однаковою для початкового і відновленого сигналів.

Для визначення похибки реконструкції проаналізовано спектри амплітуд початкового, прискореного та реконструйованого сигналів.

Проведено моделювання процедури зміни висоти голосу диктора. Порівнюючи оцінки реєстрограм вихідного та зміненого сигналу, а також оцінки їх амплітудних спектрів помічено, що вони відрізняються незначно, зате відмінності в звучанні є суттєві. Це дає можливість шифрувати ГС на приклад в системах аутентифікації користувачів. Підробити кодований сигнал майже неможливо, тому що для цього, крім оригінального ГС, знадобляться значення параметрів роботи алгоритму сповільнення/пришвидшення та зміни висоти тону.

РОЗДІЛ 4
ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Охорона праці

Основні вимоги до побудови і функціонування системи управління охороною праці (СУОП). Забезпечення функціонування та побудова СУОП в організації. Положення про СУОП, структура та зміст його розділів.

В Україні функціонує багаторівнева СУОП, функціональними ланками якої є відповідні структури державної законодавчої і виконавчої влади різних рівнів, управлінські структури підприємств і організацій, трудових колективів.

Залежно від спрямування вирішуваних завдань всі ланки СУОП можна розділити на дві групи:

• ланки, що забезпечують вирішення законодавчо-нормативних, науково-технічних, соціально-економічних та інших загальних питань охорони праці;

• ланки, до функціональних обов'язків яких входить забезпеччення безпеки праці в умовах конкретних організацій, підприємств.

До першої групи належать органи державної законодавчої ініціативи та органи державного управління охороною праці:

- Верховна Рада України;

- Кабінет Міністрів України;

- Державна служба гірничого нагляду та промислової безпеки України (Держгірпромнагляд України);

- міністерства та інші центральні органи державної виконавчої влади;

- Фонд соціального страхування від нещасних випадків і профзахворювань;

- місцева державна адміністрація, органи місцевого самоврядування.

Верховна Рада України зі своєї ініціативи у взаємодії з відповідними структурами державної виконавчої влади визначає державну політику в сфері охорони праці, вирішує питання щодо удосконалення і розвитку законодавчої бази охорони праці, соціальні питання, пов'язані зі станом умов і охорони праці.

Кабінет Міністрів України забезпечує реалізацію державної політики в сфері охорони праці, виходячи із стану охорони праці в державі, організує розробку загальнодержавних програм відповідно до поліпшення цього стану, затверджує ці програми і контролює їх виконання, визначає функції органів виконавчої влади щодо вирішення питань охорони праці і нагляду за охороною праці.

Для вирішення цих питань при Кабінеті Міністрів України функціонує Національна рада з питань безпечної життєдіяльності населення, яку очолює віце- прем'єр-міністр України.

Держгірпромнагляд України здійснює комплексне управління охороною праці на державному рівні, реалізує державну політику в цій сфері, розробляє за участі відповідних органів державної програми в сфері охорони праці, координує роботу державних органів і об'єднань підприємств із питань безпеки праці, розробляє і переглядає разом з компетентними органами систему показників і обліку умов і безпеки праці, здійснює міжнародне співробітництво з питань охорони праці і нагляд за охороною праці в державі тощо.

Рішення Держгірпромнагляду України, що відноситься до її компетенції, обов'язкові для виконання всіма міністерствами, іншими центральними органами державної виконавчої влади, місцевими державними адміністраціями, місцевими радами народних депутатів і підприємствами.

Фонд соціального страхування від нещасних випадків здійснює профілактику нещасних випадків і профзахворювань, а також координацію всієї страхової діяльності, пов’язаної з охороною праці.

Міністерство праці і соціальної політики України здійснює також державну експертизу умов праці, визначає порядок і здійснює контроль за якістю проведення атестації робочих місць згідно з їх відповідністю нормативним актам про охорону праці, бере участь у розробці нормативних документів про охорону праці.

Інші міністерства і центральні органи державної виконавчої влади як ланки системи управління охороною праці визначають науково-технічну політику галузі з питань охорони праці, розробляють і реалізують комплексні заходи щодо

поліпшення безпеки праці, здійснюють методичне керівництво діяльністю підприємств галузі з охорони праці, співробітничають з галузевими профспілками щодо вирішення питань безпеки праці, організовують у встановленому порядку навчання і перевірку знань правил і норм охорони праці керівниками і фахівцями галузі, створюють, у разі необхідності, професійні воєнізовані аварійно-рятувальні формування, здійснюють внутрішній контроль за станом охорони праці.

Для забезпечення виконання перелічених функцій в апаратах міністерств і інших центральних органів державної виконавчої влади створюються служби охорони праці.

Місцеві державні адміністрації й органи місцевого самоврядування в межах підвідомчої їм території забезпечують реалізацію державної політики в сфері охорони праці, формують за участі профспілок місцеві програми заходів щодо поліпшення безпеки, гігієни праці і виробничого середовища, здійснюють контроль за дотриманням нормативних актів про охорону праці. Для забезпечення виконання названих функцій при місцевих органах державної виконавчої влади створюються відповідні структурні підрозділи.

Управлінські структури підприємств забезпечують в умовах конкретних виробництв реалізацію вимог законодавчих і нормативних актів про охорону праці з метою створення безпечних і нешкідливих умов праці, попередження виробничого травматизму і професійних захворювань, вирішують весь комплекс питань з охорони праці, пов'язаних з даним виробництвом. У своїй діяльності стосовно охорони праці управлінські структури підприємств взаємодіють з комісією з питань охорони праці підприємства (за наявності такої), з профспілками підприємства та уповноваженими трудових колективів.

СУОП в умовах конкретної організації, на конкретному об'єкті завжди є багаторівневою системою управління, у якій верхнім рівнем є державне управління, а нижнім - управління охороною праці на конкретному об'єкті. Як проміжні рівні управління можуть виступати відомче, регіональне управління, а також управління в об'єднанні, тресті тощо.

Слід зазначити, що вихідні параметри СУОП визначаються, виходячи з вимог норм, правил, проектної документації, аналізу фактичного стану виробничої ситуації і ряду факторів виробничого середовища, тому СУОП варто віднести до категорії звичайних, багатоконтурних систем, які піддаються програмуванню. Багатоконтурність систем управління в даному випадку пояснюється складністю об'єкта управління, його великою інерційністю, складністю і інерційністю реалізації управліньських впливів.

Правовою основою СУОП є: Конституція України, Кодекс законів про працю України, Закони України «Про охорону праці» і «Про загальнообов'язкове державне соціальне страхування від нещасного випадку на виробництві і професійного захворювання, які спричинили втрату працездатності», накази і розпорядження Президента України, розпорядження і постанови Кабінету Міністрів, Держгірпромнагляд, Міністерства охорони здоров'я, Міністерства праці і соціальної політики, а також інших директивних органів України з питань охорони праці (органи Державного управління охороною праці).

Позитивна дія впровадження систем управління охороною праці (СУОП) на рівні організації як на зниження небезпек і ризиків, так і на продуктивність, нині визнана урядами, роботодавцями і працівниками.

4.2 Безпека в надзвичайних ситуаціях

Надзвичайні екологічні ситуації та екологічний ризик

Особливу роль у житті людини відіграють надзвичайні ситуації, що виникають під час стихійних лих або техногенних катастроф. Разом із соціальними та економічними збитками надзвичайні ситуації завдають також екологічної шкоди, що відображається в руйнуванні й деградації природних систем, забрудненні повітря, водойм і ґрунтів. У результаті виникають надзвичайні екологічні ситуації. Надзвичайні екологічні ситуації — ті ситуації, що виникають унаслідок раптових природних лих або техногенних аварій і супроводжуються великими збитками. Характерними особливостями цих ситуацій є велика гострота прояву, значні відхилення показників навколишнього

середовища від норми (перевищення граничнодопустимих концентрацій (ГДК) забруднювальних речовин у сотні, тисячі й навіть десятки тисяч разів); ураганні швидкості вітру; затоплення селітебних територій (населених пунктів); виникнення катастрофічних селевих потоків та ін.

Звичайно, такі відхилення тривають недовго — години, дні, десятки днів, іноді більше. Потім ступінь гостроти екологічного стану зменшується, хоча може залишатися досить високим. Отже, поняття надзвичайна екологічна ситуація та катастрофічна екологічна ситуація розрізняються тим, що перша триває порівняно недовго, але настає раптово та характеризується виключно високими відхиленнями стану навколишнього середовища від норми, а друга — досить тривала (як правило, роки), але має меншу гостроту прояву.

Надзвичайна ситуація за певних обставин може перетворитися на катастрофічну. Наприклад, ситуація у Чорнобильській зоні. Протягом майже місяця радіаційна обстановка в Чорнобилі була надзвичайною. Після спорудження саркофага викиди радіоактивних елементів різко зменшилися, але забруднення до того часу охопило великі території. Таке високе радіаційне забруднення продовжується вже понад два десятиріччя. За оцінкою спеціалістів, екологічна ситуація в Чорнобильській зоні є катастрофічною.

Таким чином, надзвичайні екологічні ситуації відображаються у порушенні нормального функціонування природних і природно-антропогенних систем, пов'язаних із раптовими природними або техногенними впливами (стихійні лиха, катастрофи, аварії), що супроводжуються соціальними, економічними та екологічними збитками і потребують для ліквідації особливих управлінських рішень. Збитки виявляються у загибелі та пораненні людей, погіршенні їх здоров'я, руйнуванні матеріальних об'єктів, структури природних і природно-антропогенних систем, втраті їх природно-ресурсного і екологічного потенціалу. Довготривала надзвичайна ситуація зумовлює формування зони екологічної катастрофи або екологічного лиха.

Надзвичайні екологічні ситуації виникають унаслідок дії трьох основних груп факторів:

— свідомого руйнування природного середовища, походження техніки, погіршення становища економічних об'єктів під час війн і диверсійних актів;

— руйнівних катастроф, які виникають у зв'язку з некомпетентними та помилковими технічними рішеннями (наприклад, Чорнобильська аварія);

— природних стихійних явищ. Той факт, що різко збільшилися їх частота та інтенсивність в останні десятиріччя, спеціалісти пов'язують з антропогенною стимуляцією, що спричинює посилення відхилень природних процесів від нормального рівня коливань.

Економічні збитки, завдані у зв'язку з несприятливими і небезпечними природними процесами та явищами, значно збільшилися. За деякими оцінками, вони зростають швидше, ніж показники світового валового продукту, тобто може бути досягнута межа просторового і технологічного розвитку виробництва за його здатністю компенсувати збитки, які збільшуються, від несприятливих і небезпечних явищ. Первинні процеси, що виникають у природному середовищі внаслідок цих факторів, посилюватимуться або послаблюватимуться залежно від природної обстановки (стійкість ландшафтів, погод ні умови, фаза коливань екосистеми тощо) і соціально-економічних умов (психологічна готовність і неготовність населення до ліквідації наслідків надзвичайної ситуації, технічна оснащеність спеціальних служб, економічні можливості та ін.). Таким чином, надзвичайні екологічні ситуації в більшості випадків мають комплексну природу.

Заходи щодо запобігання надзвичайним екологічним ситуаціям або подолання їх наслідків можна згрупувати у три класи:

— організаційні, серед яких розрізняють планувальні та оперативні;

— інженерно-технічні;

— технологічні.

Отже, заходи, спрямовані на запобігання надзвичайним екологічним ситуаціям та подолання, їх можна поділити на два типи: заходи, спрямовані на зниження піддатливості об'єктів небезпечним впливам, і заходи, спрямовані на зниження чутливості об'єктів до небезпечних впливів. У першому випадку здійснюють заходи з метою зовнішнього захисту об'єктів, виключення тих чи

інших територій з використання у виробничих цілях тощо. Зниження чутливості об'єктів до небезпечних впливів досягається, насамперед, за рахунок досконаліших технологій, шляхом регулювання технологічних режимів у зв'язку з природними циклами, створення системи дублювання об'єктів, інформаційних систем і систем швидкого реагування.

Основні функції щодо запобігання надзвичайним екологічним ситуаціям та подолання їх на державному рівні виконують міністерства з надзвичайних ситуацій.

Ризик — це об'єктивне поняття, він пов'язаний практично з будь-якою діяльністю людини. Уміння усвідомлювати ступінь ризику дає змогу людині оцінити власні можливості й вибрати напрями поведінки при цьому. Під сутністю терміна ризик розуміють імовірність, по-перше, будь-якої небезпечної події; по-друге, негативних наслідків від неї та обсягу очікуваних збитків. Одні ризики конкретні, інші — не мають такого визначення. Існують професійні ризики (наприклад, небезпека професійних захворювань) і такі, яких зазнає все населення (екологічний, економічний, геологічний, політичний ризики).

Предметом нашого дослідження є екологічний ризик, чіткого визначення якого досі немає. М.Ф. Реймерс вважає, що це ймовірність наслідків будь-яких (специфічних або випадкових, поступових або катастрофічних) антропогенних змін природних об'єктів і факторів[*22]. З екологічним ризиком пов'язані поняття екологічної безпеки і небезпеки. Ці альтернативні категорії стосуються населення як реципієнта дії навколишнього середовища за його відповідно несприятливого чи сприятливого статусу.

Екологічний ризик пов'язаний із такими групами факторів: 1) техногенними; 2) природними; 3) військовими; 4) соціально-економічними; 5) політичними; 6) тероризмом.

Техногенний екологічний ризик виникає у зв'язку з аваріями на ЛЕС, аваріями танкерів, на небезпечних хімічних виробництвах, під час руйнування гребель водосховищ тощо. Причинами аварій є інтенсивність технологічних процесів та зв'язків, висока концентрація виробництва, ресурсомісткість і

багатовідходність технологій, погана оснащеність очисними й утилізаційними пристроями.

Природний екологічний ризик пов'язаний із ймовірністю вияву багатьох несприятливих природних явищ, таких як землетруси, вулканізм, селі, повені, цунамі та ін. Потрібно враховувати особливості геологічної будови (властивості гірських порід, наявність або відсутність розламів тощо), рельєфу (наприклад, посилення ризику забруднення в улоговинах), ландшафтів (ступінь їх стійкості до техногенних навантажень). Варто також зважати на сусідство цінних та унікальних природних об'єктів, територій особливого режиму охорони. Екологічний ризик збільшується за високої густоти населення, а також залежить від характеру сприйняття населенням подій, що відбуваються. Відомо, що катастрофічні наслідки аварій і стихійних природних явищ різко зростають у результаті психологічної неготовності населення до таких подій.

Особливу групу факторів виникнення екологічного ризику становлять воєнні дії, які зумовлюють різноманітні зміни навколишнього середовища та безпосередньо впливають на людину й інші суб'єкти. Екологічний ризик пов'язаний також із соціально-економічними факторами. Йдеться про ймовірність виникнення несприятливих екологічних ситуацій у разі прийняття рішень про будівництво тих чи інших небезпечних об'єктів у зв'язку з соціальною й економічною потребами такого будівництва. До цієї категорії належить будівництво багатьох АЕС, створення небезпечних хімічних виробництв, транспортних систем. У деяких випадках аналогічні рішення пов'язані з політичними факторами.

Нині є та розробляється велика кількість науково обґрунтованих постанов, нормативів, правил, державних стандартів, за якими регламентується господарська діяльність, встановлюються граничнодопустимі концентрації шкідливих і токсичних компонентів у ґрунтах, підземних і поверхневих водах тощо. На основі цих документів та екологічного законодавства в Україні розроблено систему заходів на державному, відомчих та об'єктних рівнях, що регламентують ведення екологічно безпечної господарської діяльності, будівництво різних споруд, межі забруднення природного середовища в рамках

не лише окремих локальних систем, а й великих регіонів, держави в цілому. Такі заходи можна об'єднати у три основні групи — соціально-організаційні, оцінювально-прогнозні та технічні. Усі види заходів взаємопов'язані і є основою для організації безпечної життєдіяльності. Якщо їх правильно дотримуватися, можна не тільки зберегти стан навколишнього середовища, а й поліпшити його, уникнути екологічно небезпечних явищ і катастроф, зумовлених антропогенно-техногенною діяльністю.

4.3 Висновки до розділу

В розділі «Охорона праці та безпека в надзвичайних ситуаціях» описано основні вимоги до побудови і функціонування системи управління охороною праці, а також надзвичайні екологічні ситуації та екологічний ризик.

ВИСНОВКИ

Розглянуто особливості обміну інформацією, де основним носійним сигналом є мовні чи голосові сигнали. Показано важливість перетворення таких сигналів для збільшення швидкості обміну даними в цифрових системах зв'язку. При цьому застосовуються спеціальні пристрої – вокодери, які призначені для стиснення ГС, зокрема із врахуванням психоакустичного ефекту маскування.

Проаналізовано методи кодування ГС вокодерами та показано ефективність застосування класу фазових вокодерів, що використовують алгоритм, який може інтерполювати інформацію, наявну в частотній і часовій областях аудіосигналів, використовуючи інформацію про фазу, отриману з частотного перетворення. Комп'ютерний алгоритм дозволяє модифікувати цифровий звуковий файл у частотній області (зазвичай розширення/стиснення за часом і зсув висоти).

Проаналізувавши недоліки, які притаманні сучасним алгоритмам роботи фазових вокодерів, показано актуальність задачі удосконалення чи розроблення нових алгоритмів кодування аудіосигналів, зокрема і голосових, для побудови фазових вокодерів.

Проаналізовано відомі базові дослідження, які проводились для розроблення ефективних алгоритмів кодування аудіосигналів в фазових вокодерах, зокрема розглянуто дослідження Фланагана і Голдена, Рабінера-Шафера, Долсона, Ларош і Долсона та показано загальні вимоги та необхідні етапи обробки ГС в фазових вокодерах.

Проведено реалізацію основних етапів обробки ГС для фазових фокодерів в середовищі Matlab. При цьому виконано аудіозапис з допомогою програми Adobe Audition та стандартної комп'ютерної гарнітури і наступне опрацювання його в середовищі Matlab. Як алгоритм опрацювання використано підходи, описані в працях Ларош і Долсона, що передбачає часове стиснення/розтягування аудіосигналу, а також зміну висоти голосу диктора в аудіофайлі.

Проведено моделювання процесу стиснення ГС та наступного його розтягування і оцінено похибку, яка характеризує процес кодування в такому вокодері. Під час прослуховування виявлено, що звукова чіткість сигналу залишалася практично однаковою для початкового і відновленого сигналів.

Для визначення похибки реконструкції проаналізовано спектри амплітуд початкового, прискореного та реконструйованого сигналів.

Проведено моделювання процедури зміни висоти голосу диктора. Порівнюючи оцінки реєстрограм вихідного та зміненого сигналу, а також оцінки їх амплітудних спектрів помічено, що вони відрізняються незначно, зате відмінності в звучанні є суттєві. Це дає можливість шифрувати ГС на приклад в системах аутентифікації користувачів. Підробити кодований сигнал майже неможливо, тому що для цього, крім оригінального ГС, знадобляться значення параметрів роботи алгоритму сповільнення/пришвидшення та зміни висоти тону.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. J. L. Flanagan, R. M. Golden, "Phase Vocoder," Bell System Technical Journal, November 1966, 1493-1509. (http://www.ee.columbia.edu/~dpwe/e6820/papers/FlanG66.pdf)

2. Mark Dolson, "The phase vocoder: A tutorial," Computer Music Journal, vol. 10, no. 4, pp. 14 -- 27, 1986.

3. Jean Laroche and Mark Dolson "New Phase Vocoder Technique for Pitch-Shifting, Harmonizing and Other Exotic Effects". IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. Mohonk, New Paltz, NY. 1999. http://www.ee.columbia.edu/~dpwe/papers/LaroD99-pvoc.pdf

4. D.Ellis. A Phase Vocoder in Matlab. http://www.ee.columbia.edu/~dpwe/resources/matlab/ 2003.

5. A.Gotzen, N.Bernardini, D.Arfib. “Traditional (?) Implementations of a Phase-vocoder: the Tricks of the Trade”. Proc. Of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.

6. Dudley, H., The Vocoder. Bell Labs Record, Vol.17, 1939, pp.122-126.

7. David E.E., Schroeder M.R., Logan B.F., Prestigiacomo A.J. New Applications of Voice-Excitation to Vocoders, roc. Stockholm Speech Comm. Seminar, R.I.T., Stockholm, Sweden, September, 1962.

8. Carlson, J.P. “Digitalized Phase Vocoder”, Proc. Conf. On Speech Comm. and Proc., Boston, Mass., November 1967.

9. Математичне та комп’ютерне моделювання електрокардіосиґналів у системах голтерівського моніторинґу / Л.Є. Дедів, А.С. Сверстюк, І.Ю. Дедів, М.О. Хвостівський, В.Г. Дозорський, Є.Б. Яворська. – Львів: Видавництво «Магнолія - 2006», 2021. – 120 с. ISBN 978-617-574-218-1.

10. Математичне моделювання, методи та програмне забезпечення опрацювання дихальних шумів у комп'ютерних аускультативних діагностичних системах / І.Ю. Дедів, А.С. Сверстюк, Л.Є. Дедів, В.Г. Дозорський, М.О.

Хвостівський. – Львів: Видавництво «Магнолія - 2006», 2021. – 126 с. ISBN 978-617-574-219-8.

11. Методичні рекомендації з оформлення кваліфікаційних робіт бакалавра за спеціальністю 172 «Телекомунікації та радіотехніка» уклад.: Дунець В.Л., Хвостівський М.О. Дедів І.Ю. Тернопіль: ТНТУ імені Івана Пулюя, 2021 р. – 72с.

12. Дозорський В.Г., Дозорська О.Ф., Дедів Л.Є., Дедів І.Ю., Паньків І.М., Яворська Є.Б. Структура системи відбору біосигналів для задачі відновлення комунікативної функції людини. Вісник Хмельницького національного університету: технічні науки. – Хмельницький: редакція журналу "Вісник Хмельницького національного університету". – 2019. - №2(271) – с. 183-186.

13. Khvostivska L.V., Osukhivska H.M., Khvostivskyi M.O., Dediv S.Y. Development of methods and algorithms for a stochastic biomedical signal period calculation in medical computer diagnostic systems. Visnyk NTUU KPI Seriia - Radiotekhnika Radioaparatobuduvannia, /Категорія В/ 2019. Вип. 79. С. 78-84. doi: 10.20535/RADAP.2019.79.78-84.

14. Dozorska O., Yavorska E., Dozorskyi V., Pankiv I., Dediv L. Dediv I. The Method of Indirect Restoration of Human Communicative Function. Proc. of the 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), CADSM'2019, (pp. 19–22). Polyana-Svalyava (Zakarpattya), UKRAINE 978-1-7281-0053-1/19.

15. Дозорська О.Ф., Яворська Є.Б., Дозорський В.Г., Дедів Л.Є., Дедів І.Ю. Метод виявлення ознак основного тону в структурі електроміографічних сигналів для задачі компенсації порушеної комунікативної функції людини», Вісник НТУУ "КПІ". Серія Радіотехніка, Радіоапаратобудування, (81), с. 56-64. doi: 10.20535/RADAP.2020.81.56-64.

16. Khvostivska L., Khvostivskyy M., Dunetc V., Dediv I.. Mathematical and Algorithmic Support of Detection Useful Radiosignals in Telecommunication Networks. 2nd International Workshop on Information Technologies: Theoretical and Applied Problems, ITTAP 2022. CEUR Workshop Proceedings. Ternopil 22- 24 November 2022. Vol 3309, P. 314-318. ISSN 1613-0073.

17. Гевко О.В., Дозорський В.Г., Дедів Л.Є., Дедів І.Ю., Дозорська О.Ф. Структурний синтез вібромасажної апаратури. Перспективні технології та прилади, № 20, Луцьк, 2022. – с. 23-31.

18. Dozorskyi V., Dediv I., Sverstiuk S., Nykytyuk V., Karnaukhov A. The Method of Commands Identification to Voice Control of the Electric Wheelchair. Proceedings of the 1st International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2023). CEUR Workshop Proceedings. Ternopil, Ukraine, June 14-16, 2023. P.233-240. ISSN 1613-0073.

19. Khvostivska L., Khvostivskyi M., Dediv I., Yatskiv V., Palaniza Y. Method, Algorithm and Computer Tool for Synphase Detection of Radio Signals in Telecommunication Networks with Noises. Proceedings of the 1st International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2023). CEUR Workshop Proceedings. Ternopil, Ukraine, June 14-16, 2023. P.173-180. ISSN 1613-0073.

20. Khvostivska L., Khvostivskyi M., Dunets V., Dediv I. Mathematical, algorithmic and software support of synphase detection of radio signals in electronic communication networks with noises. Scientific Journal of TNTU (Tern.), vol 111, no 3, 2023. pp. 48–57.

21. Основи технології радіоелектронних апаратів : навчальний посібник / Р. А. Ткачук, В. Г. Дозорський, Л. Є. Дедів, І. Ю. Дедів. - Тернопіль : Тернопільський національний технічний університет імені Івана Пулюя, 2017. - 336 с.

22. Паляниця Ю.Б., Марценюк А.С., Дунець В.Л., Бучинський В.М., Паламар М.І. Дрон з блоком надвисоких частот для виявлення та знешкодження вибухових пристроїв та мін. Матеріали ІІІ Міжнародної наукової конференції молодих учених та студентів «Воєнні конфлікти та техногенні катастрофи: історичні та психологічні наслідки» Тернопільського національного технічного університету імені Івана Пулюя: зб. тез доповідей, 20-21.04.2023 р. Тернопіль: ТНТУ, 2023. С. 158-159. ISBN 978-617-7875-32-0.

23. Дунець В.Л., Хвостівська Л.В., Паляниця Ю.Б. Математичне, алгоритмічне та програмне забезпечення оцінювання завадозахищеності каналів

зв’язку з балансною модуляцією. Збірник наукових праць Вісник НУВГП, серія технічні науки, випуск 4 (104), 2023. - С. 95-107. ISSN: 2306-5478.

24. Гевко І.Б., Дунець В.Л., Паляниця Ю.Б., Марценюк А.С., Химич Г.П. Дрон з імпульсним квантовим генератором для знешкодження об’єктів ураження. Матеріали IV Міжнародної наукової конференції «Воєнні конфлікти та техногенні катастрофи: історичні та психологічні наслідки» Тернопільського національного технічного університету імені Івана Пулюя: зб. тез доповідей, 18-19.04.2024 р. Тернопіль: ТНТУ, 2024.

25. Khymych, H., Dunets, V., Duda, S., Palaniza, Y., Kornieiev, K. Dual-Polarization Yagi Antenna for Meter Wavelength Range. Radioelectronics and Communications Systems, 66(11), 2023, pp. 609–615. ISSN 0735-2727. DOI: 10.3103/S0735272722080039.

26. Lyashuk, Oleg, Mykola Stashkiv, Yurii Palianytsia, Roman Khoroshun and Dmytro Mironov. Fourier Model Fitting for Vibration Analysis of Car Wheel Suspension Enhanced Damping System. Proceedings of the 4rd International Workshop on Information Technologies: Theoretical and Applied Problems, 23-25 October 2024, Ternopil, Ukraine. ITTAP-2024, 2024. ISSN 1613-0073.

27. Паляниця Ю.Б., Сверстюк А.С., Шадріна Г.М. Математичне та комп’ютерне моделювання фонокардіосиґналів для удосконалення кардіодіагностичних систем / Ю.Б. Паляниця, А.С. Сверстюк, Г.М. Шадріна – Львів: Видавництво «Магнолія - 2006», 2020. – 106 с. ISBN 5-211-05310-9.

28. Dozorskyi V., Dediv L., Kovalyk S., Dozorska O., Dediv I. (2024) Design of the endoskeleton of a biocontrolled hand prosthesis. Scientific Journal of TNTU (Tern.), vol. 115, no 3, pp. 100-111.

29. Техноекологія та цивільна безпека. Частина «Цивільна безпека». Навчальний посібник / В.С. Стручок, – Тернопіль: ТНТУ ім. І.Пулюя, 2022. – 150 с.

30. Стручок В.С. Безпека в надзвичайних ситуаціях. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної бо та заочної (дистанційної) форм навчання / В.С.Стручок. — Тернопіль: ФОП Паляниця В. А., 2022. — 156 с.

ДОДАТКИ

Додаток А

УДК 621.395.625.6

Ірина Дедів, канд. техн. наук, доц., Маркіян Пліс, Валентин Степанов, Сергій Брегін, Володимир Лотоцький

Тернопільський національний технічний університет імені Івана Пулюя

ЗАДАЧА ОЦІНЮВАННЯ РОЗБІРЛИВОСТІ МОВИ ДЛЯ ПОКРАЩЕННЯ ЯКОСТІ СИСТЕМ ГРОМАДСЬКОГО ОПОВІЩЕННЯ

Iryna Dediv, Ph.D, Assoc. Prof., Markiyan Plis, Valentyn Stepanov, Sergii Bregin, Volodymyr Lototskyi

THE TASK OF SPEECH INTELLIGIBILITY ASSESSMENT FOR IMPROVING THE QUALITY OF PUBLIC ANNOUNCEMENT SYSTEMS

Сьогодні в Україні, в умовах війни, особливо гостро стоїть питання своєчасного оповіщення громадян про ймовірні небезпечні непередбачувані ситуації, зокрема повітряні ракетні атаки, обмеження руху та пересування через ймовірності вибухів зокрема, виникнення пожеж чи значні викиди шкідливих речовин в повітря, воду тощо. З цією метою використовуються системи громадського оповіщення, що є організованими спеціалізованими системами, які призначені для своєчасної передачі певних сигналів та інформаційних повідомлень стосовно цивільного захисту до відповідних органів влади, підприємств, організацій, установ та населення. Ці системи включають в себе технологічні підходи та способи оповіщення, спеціалізоване обладнання, прилади та канали зв'язку. Метою роботи систем оповіщення є забезпечення максимально швидкого інформування якомога більшої кількості людей про безпосередню небезпеку. Громадські (публічні) оповіщення здійснюються дистанційно шляхом використання електромеханічних сирен (зовні), мереж мовлення всіх частотних типів, систем телевізійного та мобільного зв'язку. Основна функція, яку виконує система оповіщення в аварійній ситуації, – це трансляція мовних повідомлень, спрямованих на запобігання паніці людей, та передача інформації про напрям руху для евакуації.

В поширених сьогодні варіантах побудови систем оповіщення як основний інформаційний сигнал про виникнення надзвичайної ситуації використовується мовний чи голосовий сигнал, що являє собою голосове повідомлення, яке повинне бути однозначно донесене та сприйняте людьми чи персоналом, що перебуває в зоні дії цієї системи. При цьому, часто для підвищення ефективності технічних засобів оповіщення, зокрема підвищення випромінюваної гучномовцями потужності, проводиться попередня обробка таких сигналів, зокрема фільтрація, що передбачає подавлення високочастотних та низькочастотних складових сигналу-повідомлення та підсилення середньочастотних складових. Відповідно, часто системи оповіщення транслюють готосові повідомлення в ділянці саме середніх частот голосового сигналу. Але це в свою чергу разом із впливом зовнішніх факторів значно впливатиме на якість того голосового повідомлення, яке будуть чути люди, для яких воно транслюється. Ці і додаткові фактори значно спотворюватимуть сигнали повідомлення та вливатимуть на степінь сприйняття оточуючими людьми таких повідомлень від системи оповіщення.

В цьому плані важливим є контроль якості роботи системи з можливістю коригування параметрів чи характеристик голосових повідомлень для забезпечення надійного та однозначного їхнього сприйняття оточуючими людьми. Власне для цього і потрібно розробити новий чи обґрунтувати вибір відомого методу оцінювання розбірливості мови.