

ЗАХИСТ МОВНОЇ ІНФОРМАЦІЇ МЕТОДАМИ КОДУВАННЯ З АДАПТИВНИМ ПЕРЕДБАЧЕННЯМ

У статті розглядаються основні принципи побудови схем кодування мовної інформації методами адаптивного передбачення. Наводяться структурні схеми кодерів і декодерів, а також результати оцінки якості різних схем.

Ключові слова: мовна інформація, адаптивне передбачення, вокодер, форманти, диференціальна імпульсно-кодова модуляція.

При передачі інформації в стільникових мережах зв'язку і передачі мови в мережі Інтернет (ІР-телефонія) широке застосування знайшли методи кодування з передбаченням. Принцип передбачення полягає в наступному (рис. 1): на передавальній стороні встановлюється передбачувач, який за отриманою в попередній момент інформацією передбачає подальшу інформацію. При надходженні інформації від передавача передбачене і істинне значення віднімаються і в канал зв'язку передається помилка передбачення. На приймальному кінці встановлюється передбачувач, який за попередньою інформацією виробляє послідовні значення сигналу, — ті ж самі, що і передбачував передавач, — і коригує (підсумовує) з сигналом помилки, що поступив. Передача різницевого сигналу у більшості випадків вимагає меншого числа розрядів.

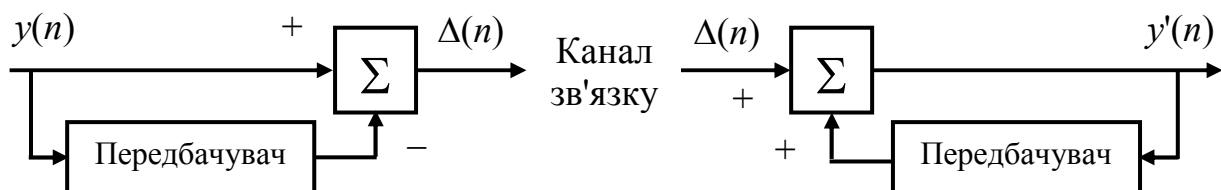


Рис. 1. Модель адаптивного передбачення

Кодування з адаптивним передбаченням використовується для передбачення форми сигналу, яка має досить прогнозовані ділянки.

Всяке кодування з передбаченням ґрунтується на апроксимації форми сигналу, тобто кодування кривої за допомогою значень цієї кривої в окремих точках і відновлення її форми на приймальному кінці по цих окремих відліках. У техніці кодування мови широке поширення отримала лінійна апроксимація. При цьому способі кожна крива представляється сумою величин попередніх відліків, помножених на коефіцієнти, що називаються коефіцієнтами передбачення:

$$y(T) = \sum_{t=1}^p a_t \cdot y(T-t), \quad (1)$$

де $y(T)$ — значення відліку в наступний момент часу; a_t — коефіцієнти передбачення; p — порядок моделі.

Чим вище порядок моделі, тим точніше вона наближається до істинного значення форми кривої.

Передбачення може також ґрунтуватися на різних принципах.

Простий принцип передбачення — збереження попереднього значення. Приведена формула перетвориться в $y(T) = a_1 y(T-1)$, $a_1 = 1$. Тоді в канал зв'язку передається різниця між поточним і попереднім значеннями:

$$\Delta y = y(T) - y(T-1). \quad (2)$$

Тут $y(T)$ — це апроксимоване значення сигналу.

Відновлення сигналу на прийомі полягає в тому, що до попереднього значення сигналу додається різниця: $y(T) = y(T-1) + \Delta y$.

Цей принцип лежить в основі схем диференційної імпульсно-кодової модуляції (ДІКМ), одна з можливих реалізацій якої приведена на рис. 2.

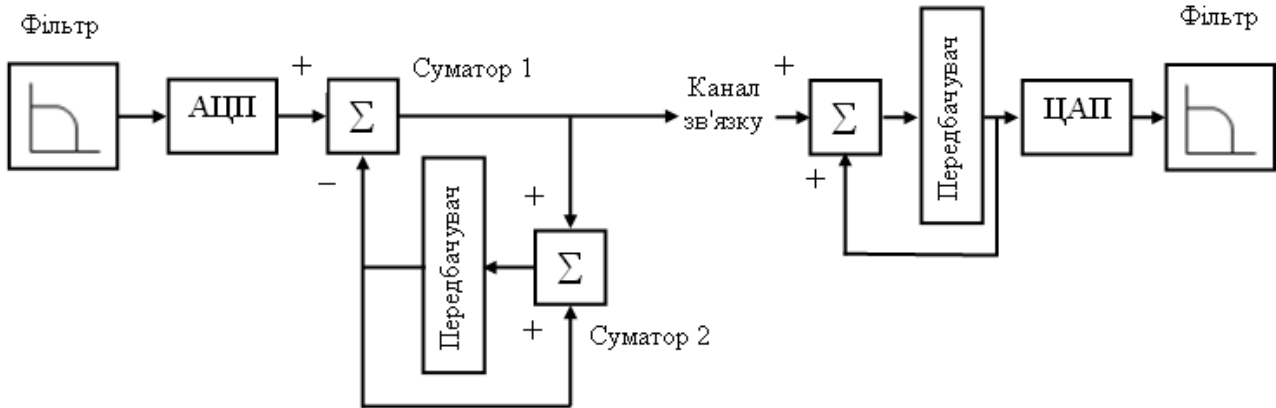


Рис. 2. Реалізація ДІКМ

Предбачувач є регістром, який накопичує значення попереднього відліку. Для визначення наступного різницевого відліку сигнал відновлюється в суматорі складанням цього відліку з попереднім значенням сигналу в суматорі 2.

З метою підвищення точності апроксимації може братися декілька відліків. При цьому подальше значення $y(T)$ визначається по формулі (1).

Ефективність кодування оцінюється виграшем від кодування і вимірюється відношенням потужності сигналу, представленого кодами рівномірного квантування, і потужністю сигналів кодування різницевого сигналу. При ДІКМ такий вигреш досягає 5 дБ. Проте система з різницевим передбаченням стає неефективною при великому значенні різницевого сигналу.

Втрата або спотворення значення різницевих відліків призводять до повного спотворення відновлюваних значень, оскільки помилка в попередньому значенні сигналу викликає катастрофічні помилки у визначенні подальших. Це вимагає передачі на приймач величини виграшу (для контролю якості сигналу, що приймається) і періодичного оновлення значень коефіцієнтів (при стійкій зміні характеру сигналу), що значно ускладнює реалізацію як кодера, так і декодера.

Одним з перспективних напрямів є методи, що засновані на імовірнісних методах прогнозування. При цьому послідовні k відліків розглядаються як k випадкових величин, і визначається вірогідність їх спільної появи — $p(x_1, \dots, x_k)$. Прогнозування усього відліку засноване на визначенні математичного очікування

$$M[y(T)] = \sum_{k=1}^p p_k \cdot y(T-t), \quad (3)$$

де p_k — вірогідність появи k -го відліку.

У канал зв'язку передається різниця між реальною величиною у момент t і його прогнозованим значенням (математичним очікуванням): $Y(T) - M[y(T)]$.

Імовірнісні формули передбачення можуть мати найрізноманітніший характер: від формул, що припускають наявність імовірнісного закону, до лінійно-шматкової апроксимації значень вірогідності. У усіх випадках приймачу передаються параметри формули прогнозування. Моментами передання можуть бути моменти зміни на основі нової статистики (адаптація параметрів).

Таким чином, ефективність кодування з адаптивним передбаченням залежить від складності адаптивної логіки і числа відліків для наступного передбачення.

Оскільки при даному методі передачі кодується різницевий сигнал, що має меншу потужність, то такий метод забезпечує високий рівень виграшу передбачення (від 13 до 20 дБ).

Розглянемо, як цей метод може бути реалізований у вокодері. Робота вокодера заснована на аналізі характерних особливостей людської мови (рис. 3).

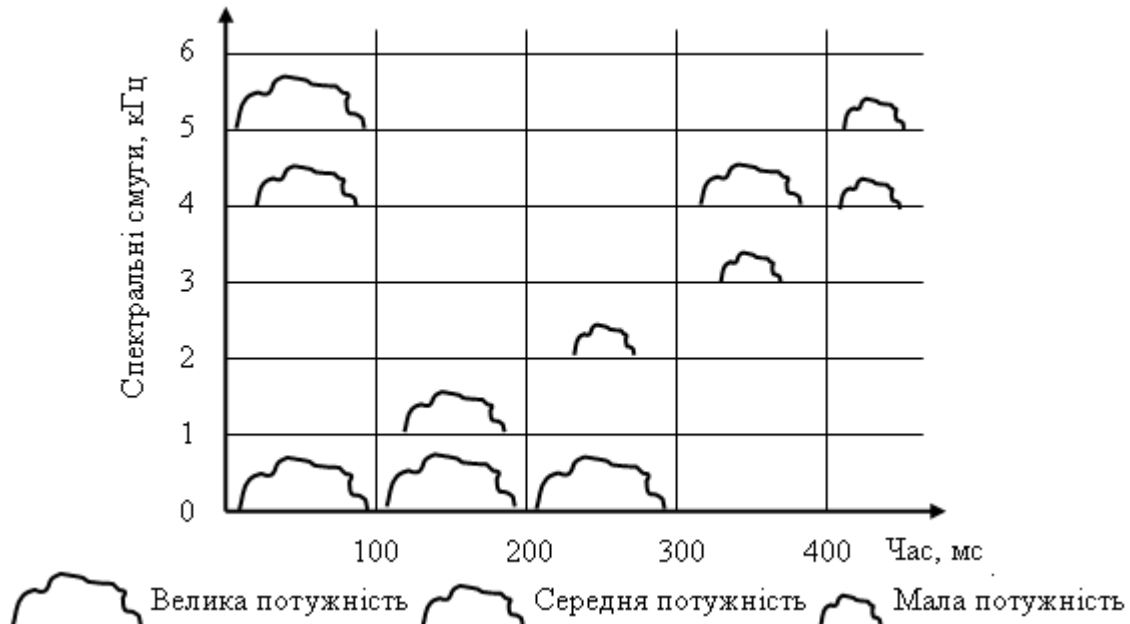


Рис. 3. Приклад розподілу енергії в частотних діапазонах

Енергія мовного сигналу розподіляється в часі тільки в деяких частотних діапазонах і розрізняється за величиною. Окремі піки енергії, що виникають в одному частотному діапазоні, називаються формантами. Звуки можуть відрізнятися не лише частотним діапазоном, але і структурою. Для них характерні піки енергії в певних частотних діапазонах і провали в інших. Частоти, на яких в даний момент виникають комбінації піків, називаються частотами формант. Гласні і дзвінкі приголосні звуки мови містять 3-4 форманти. Вивчення зразків мови показує, що в українській мові міститься 38 фонем: це 6 гласних звуків і інші приголосні. Для кодування їх номерів при передачі по цифрових каналах досить 6 біт.

Коли людина говорить, вона проводить спектрально-тимчасову модуляцію широкосмугового сигналу. Корисна інформація міститься тільки в інтонації (зміні частоти основного тону) і в зміні спектру з тонального на шумовий і навпаки. Лінійна модель мовотворення представляє мову як систему, що складається з генератора збудження (генераторна функція) і лінійної системи з параметрами (фільтрова функція), що повільно змінюються. Для економічної передачі і зберігання мови потрібно визначити параметри генераторною і фільтровою функцій. У генераторній функції змінюється частота і амплітуда основного тону (висота і гучність голосу) і відбувається зміна виду функції (основний тон або шум). У фільтрової функції відбувається постійна зміна коефіцієнта передачі, що проявляється в зміні спектру.

Розглянемо принципи побудови систем, які моделюють людську мову, використовуючи при цьому методи цифрового кодування. Вони називаються вокодерами.

За принципом визначення параметрів фільтрової функції розрізняють наступні типи вокодерів :

- каналні (смугові);

- формантні;
- ортогональні;
- вокодері з лінійним передбаченням.

Основне завдання процесу кодування в смуговому вокодері — визначити спектр сигналу, потужності в кожному діапазоні частот за досить довгий відрізок часу, в який існує форманта. На передавальній стороні безперервний мовний сигнал перетворюється в цифровий за допомогою АЦП і подається на ряд цифрових фільтрів, кожен з яких виділяє вузьку смугу. Далі використовуються пристрої, що вимірюють і кодують значення потужності спектру в кожному діапазоні частот. На додаток до інформації про спектр вокодері визначають характер збудження (гласний або дзвінкий згідний, на відміну від глухого приголосного) і частоту основного тону для голосних або дзвінких приголосних звуків.

Отримана інформація передається на приймач, де вона використовується для управління цифровим генератором. Він представляє пам'ять, де зберігаються тимчасові відліки частот, з яких необхідно вибрати потрібну по частоті і потужності. Збудження голосних відбувається за допомогою генератора імпульсів, який відкриває на певний час генерацію основного тону. Збудження глухих звуків імітується шумовим генератором.

Ортогональні вокодері відрізняються від смугових тим, що функції фільтрів виконуються за допомогою цифрових методів (швидке перетворення Фур'є, ортогональні функції Уолша та інш.).

Формантний вокодер визначає і передає положення піку енергії в частотному діапазоні, амплітуду спектральних піків. Внаслідок цього знижується об'єм інформації, що передається. Якість відновленої мови залежить від точності визначення цих параметрів. Принцип облаштування формантного вокодера заснований на розподілі спектру на смуги і у визначенні в смугах необхідних характеристик. Але для передачі відбираються тільки дані про збуджені спектри. Це знижує вимоги до об'єму інформації, що передається. Декодер відновлює сигнал також за допомогою генерації основного тону і різних типів сигналів (шумових і імпульсних). Такий тип вокодера забезпечує передачу мови зі швидкістю до 1 Кбіт/с, але застосовується порівняно рідко із-за великих труднощів, пов'язаних з точним обчисленням формант.

Вокодер з лінійним передбаченням (LPC - Linear Predictive Coding), на відміну від попередніх типів, для передачі мови застосовує не фільтри, а систему лінійного передбачення (рис. 4 – 5).

Пряме використання передбачення дозволяє відтворювати звук, але з поганою якістю. Тому цей метод має багато різних різновидів, що покращують якість. Ці методи торкаються поліпшення параметрів збудження генераторів на приймальному кінці. Тому з трьох складових системи з передбаченням — апроксимації, передбачення і методів відновлення (збудження генераторів) мови — усі удосконалення методу лінійного передбачення торкаються останньої складової. Тому вони іноді називаються гібридними кодерами.

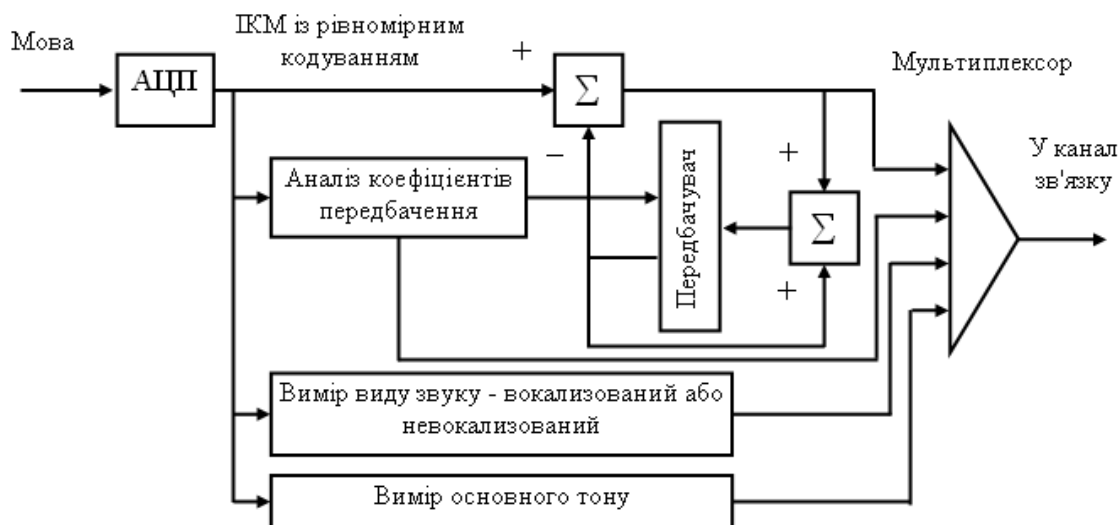


Рис. 4. Схема передавальної частини вокодера з лінійним передбаченням

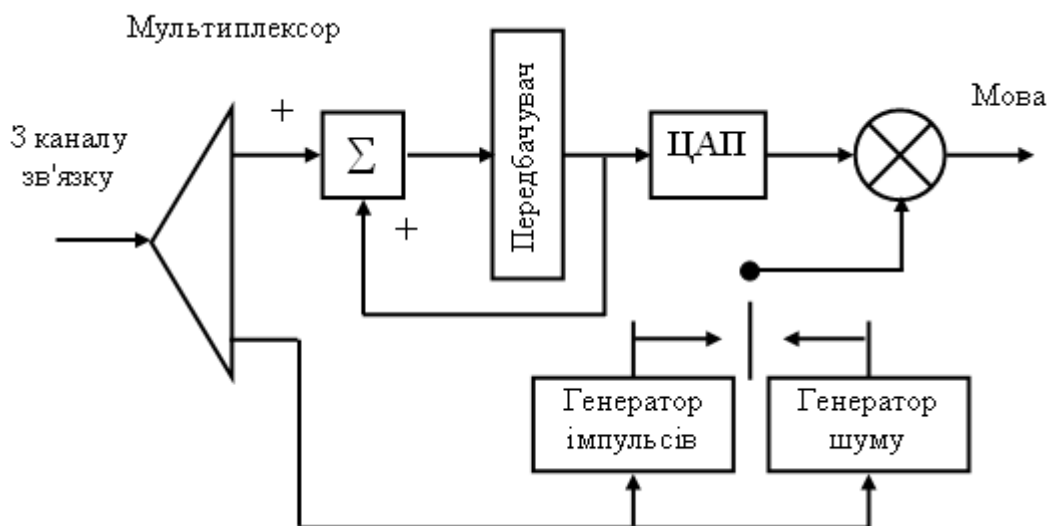


Рис. 5. Схема приймальної частини вокодера з лінійним передбаченням

Багатоімпульсне кодування (MPLPC — Multi — Pulse LPC) відрізняється від LPC тим, що передбачення торкається не основного тону, а параметрів імпульсів, що приймаються, що більше схоже на методи адаптивного кодування. На відміну від останніх прогнозується послідовність імпульсів. Такий тип кодування застосовується в система Skyphone і дозволяє передавати мову із швидкістю 9,6 Кбіт/с.

Лінійне передбачення зі збудженням усіченого залишкового сигналу (RELP LTP — Residual Excited Linear Prediction Long Term Prediction) відрізняється від попереднього тим, що в результаті обробки кодується і передбачається нижня частина мовного спектру, і це зменшує число оброблюваних відліків, що передбачаються. Застосування RELP в деяких Європейських мобільних системах дозволяє понизити швидкість передачі в каналі з 16 до 9,6 Кбіт/с без істотного зниження якості мови.

Лінійне передбачення з кодовим збудженням (CELP — Code Excited Linear Prediction) займає проміжне місце між кодерами форми сигналу і параметричними вокодерами. Аналіз параметрів мови здійснюється на інтервалах 10-30 мс, що дозволяє ефективно застосовувати CELP при швидкостях передачі від 4 до 16 Кбіт/с. Замість кодування сигналів відлік за відліком кодером різницевого сигналу застосовується кодова книга збудження — кожному різницевому сигналу зіставляється рядок сигналів, який містить набір відліків, що

відповідають передаваній залишковій послідовності на основі отриманого значення помилки. У приймачі замість декодера різницевого сигналу також застосовується кодова книга збудження.

Існує декілька різновидів кодових книг, які класифікуються:

- за принципом пошуку кодів векторів (з повним перебором, двійковий або послідовний пошук та інш.);
- по способах навчання (фіксована книга або така, що адаптується);
- по виду інформації, що зберігається (вибірки мови або реалізація шуму).

Алгоритм CELP дозволяє відтворювати мову з високою якістю. Проте він вимагає великих обчислювальних ресурсів, тому на його основі побудовано багато різновидів кодерів.

Багато з технологій, що використовують методи передбачання і вокодерні принципи перетворення, стандартизовані ІТУ-Т.

Нижче наводиться таблиця. 1, яка дозволяє порівняти якість методів при різних типах кодерів. У таблиці 2 наводяться результати оцінки різних кодерів по різних методах оцінок.

Таблиця 1. Експертна оцінка за шкалою DRT
(Diagnostic Rhyme Test – діагностичний римований тест)

Оцінка в %	Показник якості
95-100	Чудово
87-95	Добре
79-87	Задовільно
70-79	Погано
Менше 70	Незадовільно

Таблиця 2. Порівняння якості мови для основних типів кодерів

Кодер	Швидкість кодування, кбит/з	Метод оцінки		
		Діагностичний римований тест	Діагностичний критерій придатності	метод експертних оцінок
ІКМ	64	95	73	4,3
АДИКМ	32	94		4,1
LD - CELP	16	94		4,0
RPE - LTP (GSM)	13			3,5
MPLPC (Skyphone)	9,6			3,4
QCELP	9,6			3,4
CELP	8	93	68	3,7
LPC	2,4	90	54	2,5

Література

1. Берлин А.Н. Терминалы и основные технологии обмена информацией. – М.: БИНОМ. Лаборатория знаний, Интернет-университет информационных технологий - ИНТУИТ.ру, 2007.
2. Беллами Дж. Цифровая телефония. – М.: Эко-Трендз, 2004.
3. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка передачи речи. – М.: Радио и Связь, 2000.

