УДК 004.934

Михайло Лобур, Фаді Аль Келані
Національний університет "Львівська політехніка",
кафедра систем автоматизованого проектування

АНАЛІЗ МЕТОДІВ КОМПРЕСІЇ МОВНИХ СИГНАЛІВ

Подано аналіз ефективності методів компресії і визначено оптимальні методи з точки зору простоти реалізації або рівня стиску мовних сигналів.

In this paper the speech signals compression methods efficiency analysis is represented. The optimal methods for simplicity realization or compression rate for these signals is determined.

Вступ

Методи компресії мовних сигналів можна розділити на дві загальні групи: ті, що компресують без втрати якості (переважно ґрунтуються на усуненні послідовності однакових фрагментів), і ті, що компресують з частковою втратою якості сигналу; вони і набули найбільшого застосування із-за своєї гнучкості і можливості домагатися великого ступеня компресії за достатньої якості.

1. Компресія в часовій області

1.1. Нерівномірний розподіл рівнів квантування

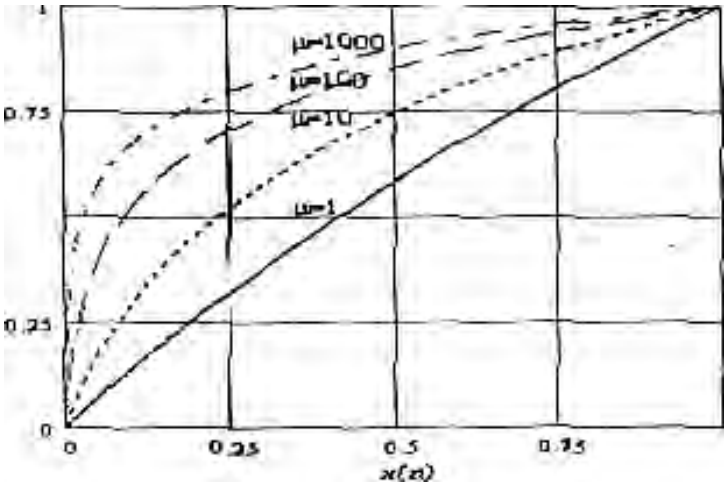
За рівномірного розподілу рівнів квантування не враховується специфіка мовного сигналу, а саме – наявність більшої кількості малих рівнів сигналу, чим великих.

Для зменшення кількості бітів для кодування вибірки мовного сигналу використовується логарифмічна залежність рівнів квантування, що називається μ-законом компандування [2]

$$y(n) = F[x(n)] = X_{\max} \cdot \frac{\log\left[1 + \mu \cdot \frac{|x(n)|}{X_{\max}}\right]}{\log[1+\mu]} \cdot sign[x(n)], \qquad (1)$$

де μ – коефіцієнт, що задає ступінь нелінійності передавальної характеристики (див. рисунок), $X_{\max}$ – максимальне значення амплітуди сигналу.

На рисунку показано сімейство передавальних характеристик для μ-закону компандування.



Сімейство передавальних характеристик для μ-закону компандування

Використання системи компандер-експандер з μ-законом дає змогу отримати постійну залежність відстотка шуму від дисперсії [2].

З дослідів визначено, що при використанні 7 біт на вибірку, відношення сигнал/спотворення дорівнює 34 дБ і досягається в широкому діапазоні частот. За рівномірного розподілу для одержання такої якості потрібне використання 11 біт.

Оптимальний вибір рівнів квантування проводиться за імовірнісним розподілом сигналу і ґрунтується на рівномірному розподілі імовірності попадання сигналу в кожний інтервал між рівнями квантування. За використання розподілу Лапласа відношення сигнал/похибка для 8 і 16-рівневого квантувача дорівнює 12.61 і 18.12 дБ, відповідно (3- і 4-бітове подання).

1.2. Адаптивне квантування

Існує два класи схем адаптивного квантування: в першому амплітуда чи дисперсія оцінюється за вхідним сигналом, а в другому – за вихідним. Розглянемо перший клас.

У цьому разі вибірка кодується кодовим словом c(n) і кроком квантування, хоча можлива передача кроку відразу, на певну кількість вибірок (сегмент). Адаптація кроку квантування дає збільшення відношення с/сп на 5.6 дБ порівняно з неадаптивним нерівномірним квантуванням.

Квантувач другого типу з адаптацією за виходом, в якого дисперсія вхідного сигналу оцінюється вже за квантованою попередньою вихідною вибіркою чи за певною кількістю вихідних вибірок.

Оцінка миттєвого значення дисперсії в цій системі можлива лише на основі попередніх відліків, оскільки вихідна вибірка одержується тільки після квантування з адаптованим кроком. За цієї побудови компресора зменшується потрібна швидкість передачі інформації, але збільшується чутливість до помилок у каналі зв'язку, які приводять до повного розузгодження передавальної і приймальної систем. Якість компресії, порівняно з адаптацією за входом є майже однаковою [I].

1.3. Різницеве квантування з лінійним передбаченням

Ґрунтуючись на кореляційній залежності між собою вибірок сигналу можна кодувати лише різницю, яка існує між сусідніми вибірками, – це найпростіший випадок, або ж кодувати різницю між вибіркою і передбаченим за попередніми вибірками значенням сигналу – на цьому і будується різницеве квантування.

В системі різницевого квантування з лінійним передбаченням використовується математична модель мовного сигналу, в якій загальний спектр, обумовлений випромінюванням, мовним трактом і збудженням, записується за допомогою лінійної системи із змінними параметрами, яку можна подати передавальною функцією [2]

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k \cdot z^{-k}}, \tag{2}$$

де G – коефіцієнт підсилення; a_k – коефіцієнти цифрового фільтра; p – порядок цифрового фільтра.

Така система збуджується імпульсною послідовністю для вокалізованих звуків і білим шумом – для невокалізованих. Параметри цифрового фільтра змінюються у часі.

При синтезі мови вибірки утвореного сигналу s(n) і сигналу збудження u(n) пов'язані простою різницевою рівністю

$$s(n) = \sum_{k=1}^{p} a_k \cdot s(n-k) + G \cdot u(n). \tag{3}$$

Лінійний передбачувач з коефіцієнтами α_k визначається як система, на виході якої

$$s(n) = \sum_{k=1}^{p} \alpha_k \cdot s(n-k). \tag{4}$$

Системна функція передбачувача **p**-го порядку являє собою поліном вигляду

$$p(z) = \sum_{k=1}^{p} \alpha_k \cdot z^{-k} . \quad (5)$$

Порівнявши формули (3) і (4), видно, що при $\alpha_k = a_k$ сигнал помилки дорівнює

$$e(n) = G \cdot u(n) . \quad (6)$$

Отже, фільтр передбачення H(z) є оберненим фільтром до системи синтезу A(z)

$$H(z) = \frac{G}{A(z)} . \quad (7)$$

Основна задача аналізу на основі лінійного передбачення ґрунтується на безпосередньому визначенні параметрів $\{\alpha_k\}$, знаходження яких зводиться до знаходження коренів лінійної системи рівнянь, що і приводить до одержання параметрів передбачення.

Використавши лінійне передбачення, можна одержати “хороше сприйняття мови за швидкостей передачі 1000–2000 біт/сек” .

1.4. Лінійне передбачення з кодовим збудженням

Цей метод дає змогу ще більше знижувати швидкість передачі при збереженні достатньої якості, що досягається за рахунок значного ускладнення обробки. Такий метод використовує бази "образів" фрагментів мовного сигналу і проводить пошук "образу", який якнайкраще відповідає аналізованому відрізку сигналу.

Сегмент сигналу проходить через частотне перетворення, за яким визначається багато частотних характеристик, таких як основний тон і його інтенсивність. Одержані характеристики використовуються для амплітудного і частотного масштабування "образу" мовного сигналу, вибраного з таблиці кодових сигналів, а також для управління синтезуючим фільтром. Множина коефіцієнтів синтезованого фільтра визначається розв'язком системи лінійних рівнянь, одержаних з критерію мінімізації середньоквадратичної помилки передбачення. Процедура пошуку "образів" і адаптація синтезованого фільтра продовжується до того часу, поки не буде досягнуто потрібної подібності синтезованого і реального сигналу.

Великою перешкодою на шляху використання цього методу є наявність багатьох зворотних зв'язків, які підвищують вимоги до апаратної реалізації, що відбивається на собівартості систем зв'язку.

2. Компресія з перетворенням сигналу

2.1. Фазовий вокодер

Дискретне перетворення Фур'є надає можливість визначити миттєвий спектр сигналу (дискретне перетворення Фур'є сегмента сигналу), передача якого і закладена в основу фазового вокодера.

Вибравши із сигналу сегмент вибірок, знаходимо його миттєвий спектр (амплітуду і фазу).

Амплітуду і похідну фази проріджують в часовій області, припускаючи про неможливість різкої зміни миттєвого спектра сигналу наступного сегмента від попереднього, кодують і передають на приймальну сторону, де проводиться відновлення (інтерполяція) виключених коефіцієнтів розкладу з подальшим переходом в часову область за допомогою оберненого перетворення Фур'є.

2.2. Смуговий вокодер

Смуговий вокодер, винайдений Дадлі [2], використовує також перетворення Фур'є, але з великим наближенням. Весь діапазон частот поділений на смуги смуговими фільтрами і передається лише значення амплітуди сигналу після кожного смугового фільтра, а також такі частотні параметри, як період основного тону і ознака тон-шум.

На приймальній стороні синтезується сигнал за допомогою спрощеного синтезатора мови, в якого система мовного тракту будується на основі переданих амплітуд сигналів після кожного смугового фільтра (подібно до керованого еквалайзера в акустиці).

Смуговий вокодер тісно пов'язаний із параметрами мовного сигналу, такими як вокалізованість чи невокалізованість звуку та період основного тону. Ці вокодери забезпечують значне зниження швидкості передачі із відповідним збільшенням спотворень. Переважно вони працюють на швидкостях 1200–9600 біт/сек [1].

2.3. Гомоморфний вокодер

Гомоморфну до згортки систему можна подати рівнянням

$$D[x(n)] = D[x_1(n) * x_2(n)] = D[x_1(n)] + D[x_2(n)] = \tilde{x}_1(n) + \tilde{x}_2(n) = \tilde{x}(n), \quad (8)$$

і навпаки

$$D^{-1}[\tilde{x}(n)] = D^{-1}[\tilde{x}_1(n) + \tilde{x}_2(n)] = D^{-1}[\tilde{x}_1(n)] * D^{-1}[\tilde{x}_2(n)] = x_1(n) * x_2(n) = x(n). \quad (9)$$

В площині Z-перетворення згортка переходить в добуток

$$X(z) = X_1(z) \cdot X_2(z). \quad (10)$$

Один з підходів синтезу такої системи є використання логарифму

$$\tilde{X}(z) = \log(X(z)) = \log(X_1(z) \cdot X_2(z)) = \log(X_1(z)) + \log(X_2(z)). \quad (11)$$

На основі гомоморфності системи (11) введено поняття про кепстр сигналу, який обраховується за формулою

$$c(n) = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \log\left[X(e^{j \cdot \omega})\right] \cdot e^{j \cdot \omega \cdot n} d\omega, \quad (12)$$

де $X(e^{j\omega})$ – спектральне подання сигналу.

Кепстр являє собою обернене перетворення Фур'є від логарифму модуля спектра. Дослідження кепстра [2] показали зручність подання кепстром сигналу, де основні параметри сигналу віддалені один від одного. Інформація про сигнал збудження знаходиться в області великих часів, а інформація про мовний тракт і форму імпульса збудження – в області малих часів. Пряме перетворення частини кепстра в області малих часів приводить до логарифму передавальної функції, що описує сумісний вплив мовного тракту, форми імпульсу збудження і випромінювання. На основі передачі кепстра в області малих часів побудовано томоморфний вокодер.

Гомоморфний вокодер з 26 значеннями кепстра, квантованими 50 раз за секунду, дає "дуже високу якість і натуральність мовного сигналу".

2.4. Компресія з використанням довільних базисів ортогональних функцій

Кожний сегмент цифрового сигналу можна подати коефіцієнтами розкладу в дискретному базисі ортогональних функцій без жодної втрати форми і параметрів сигналу, що дає можливість використовувати гнучкіші методи компресії, а також збільшити якість компресованих сигналів.

Функції є ортогональними за виконання умови

$$\int_{-\infty}^{\infty} \varphi_i(t) \cdot \varphi_j(t) dt = 0, i \neq j; j, i = 1 \cdots N. \quad (13)$$

За використання цифрового подання умова ортогональності має вигляд

$$\sum_{n=-\infty}^{\infty} \varphi_i(n) \cdot \varphi_j(n) = 0, i \neq j; i, j = 1 \cdots N. \quad (14)$$

Прямий дискретний розклад в ортогональному базисі записується

$$F(m) = \sum_{n=1}^{N} x(n) \cdot A(m,n), \tag{15}$$

де N – кількість вибірок у сегменті перетворення; A(m,n) – ядро прямого перетворення (матриця NxN).

Обернене перетворення має вигляд

$$\tilde{x}(n) = \sum_{m=1}^{N} F(m) \cdot B(n,m), \tag{16}$$

де B(n,m) – ядро оберненого перетворення (матриця NxN).

В матричному поданні пряме і зворотне перетворення записуються як

$$\begin{aligned} f &= A \cdot x; \\ \tilde{x} &= B \cdot f. \end{aligned} \tag{17}$$

Для унітарних перетворень матриця $B = A^{-1} = A^{T}$. Матриці A і B називають операторами прямого і зворотного перетворення.

В системах передачі мовного сигналу з компресією на основі розкладу в ортогональному базисі функцій повністю усувається сукупність мінімальних коефіцієнтів розкладу сегмента сигналу, якими можна знехтувати або передавати з меншою точністю.

На приймальній стороні може бути відсутнім відновлення втрачених коефіцієнтів розкладу, при цьому вони прирівнюються до нуля, але в такому разі зменшується загальна якість компресованого сигналу.

В математиці відома досить велика кількість ортогональних перетворень таких як, наприклад, Фур'є, Уолша, синусне, косинусне, нахилене тощо.

Вибравши відповідне перетворення, можна досягти різних результатів як компресії, так і якості самого компресованого сигналу. Від цього залежить форма спотворень, що виникають при компресії.

Висновок

В роботі отримано такі результати: найбільший рівень стиску мовних сигналів при використанні як міри віддалі між сигналами поняття відношення сигнал/шум або коефіцієнт кореляції одержується для перетворення Карунена–Лоєва. Найпростіше і найшвидкісніше є косинусне перетворення. Максимальний рівень стиску, що одержується для перетворення Карунена–Лоєва, пояснюється тим, що для цього перетворення характерна найменша серед всіх ортогональних перетворень дисперсія коефіцієнтів розкладу. Проте недолік цього перетворення пов’язаний з відсутністю швидких алгоритмів для цього перетворення.

1. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передача речевых сигналов. – М., 1985. 2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов / Пер. с англ.; Под ред. М.В. Назарова, Ю.В. Прохорова. – М., 1981. 3. Орищенко В.И. И др. Сжатие данных в системах сбора и передачи информации. – М., 1995. 4. Прохоров Ю. Статистические модели и рекурентное представление радиосигналов. – М., 1984. 5. Свириденко В.А. Анализ систем со сжатием данных. – М., 1977. 6. Свириденко В.А. Передача сообщений с повышенной информативностью. – М., 1983. 7. Претт У. Методы передачи изображений. Сокращение избыточности / Пер. с англ. – М., 1983. 8. Претт У. Цифровая обработка изображений. Т. 2 / Пер. с англ. – М., 1982.