

## Лабораторная работа № 1

### Кодирование речевых сигналов на основе линейного предсказания

Основной принцип метода линейного предсказания состоит в том, что текущий отсчет речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отсчетов. Коэффициенты предсказания – это весовые коэффициенты, используемые в линейной комбинации. Коэффициенты предсказания определяются однозначно из условия минимизации среднего квадрата разности между отсчетами речевого сигнала и их предсказанными значениями (на конечном интервале).

Основные положения метода линейного предсказания хорошо согласуются с моделью речеобразования, где речевой сигнал представляется в виде сигнала на выходе линейной системы с переменными во времени параметрами, возбуждаемой квазипериодическими импульсами (в пределах вокализованного сегмента) или случайным шумом (на невокализованном сегменте). Метод линейного предсказания позволяет точно и надежно оценить параметры этой линейной системы с переменными коэффициентами.

Рассмотрим модель речеобразования, показанную на рис.1.1. Эта система возбуждается импульсной последовательностью для вокализованных звуков речи и шумом для невокализованных. Таким образом, модель имеет следующие параметры: классификатор вокализованных и невокализованных звуков, период основного тона для вокализованных звуков, коэффициент усиления  $g$ , и коэффициенты  $\{a_k\}$  цифрового фильтра. Все эти параметры медленно меняются во времени. Остановимся подробно на задаче определения коэффициентов цифрового фильтра  $\{a_k\}$ .

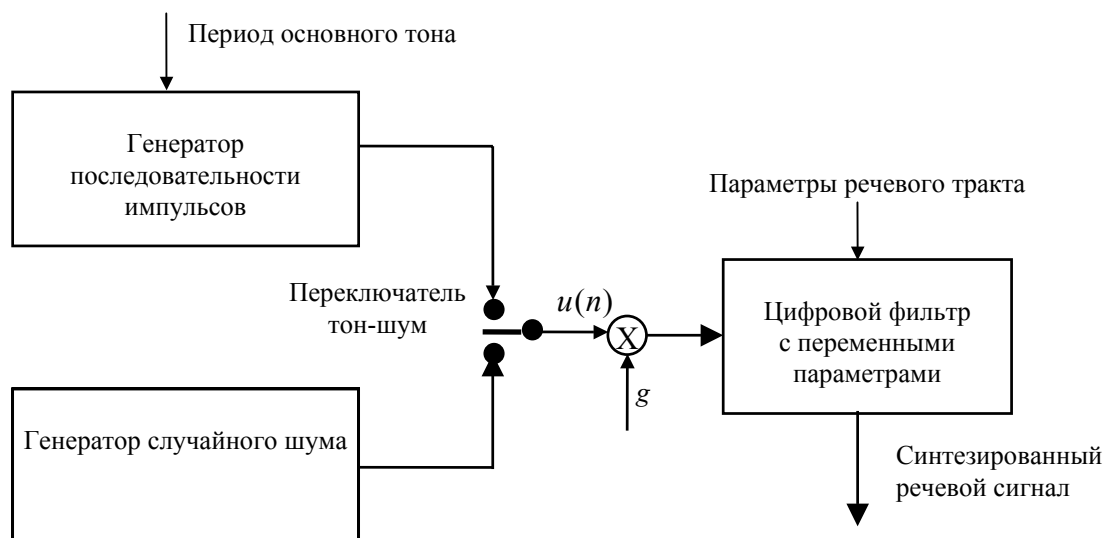


Рис.1.1. Структурная схема упрощенной модели речеобразования

Предположим, что отсчеты речевого сигнала  $s(n)$  связаны с сигналами возбуждения  $u(n)$  простым разностным уравнением

$$s(n) = \sum_{k=1}^p a_k s(n-k) + gu(n). \quad (1.1)$$

В этом случае передаточная функция линейной системы с входом  $u(n)$  и выходом  $s(n)$  имеет вид

$$H(z) = S(z)/U(z) = g / (1 - \sum_{k=1}^p a_k z^{-k}),$$

где  $z$  – формальная переменная;  $S(z)$  и  $U(z)$  –  $z$ -преобразования речевого сигнала  $s(n)$  и сигнала возбуждения  $u(n)$ .

Линейный предсказатель с коэффициентами  $a_k$  определяется как система, на выходе которой в момент времени  $n$  имеем

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k). \quad (1.2)$$

Системная функция предсказателя  $p$  – го порядка представляет собой полином вида

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

Погрешность предсказания определяется как

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

Другими словами, погрешность предсказания представляет собой сигнал на выходе системы с передаточной функцией

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

Таким образом, если сигнал точно удовлетворяет модели (1.1) и  $a_k = \alpha_k$ , то  $e(n) = gu(n)$ . Отсюда следует, что фильтр погрешности предсказания  $A(z)$  является обратным фильтром для системы с передаточной функцией  $H(z)$ , соответствующей уравнению (1.1), т.е.  $H(z) = g / A(z)$ .

Основная задача анализа на основе линейного предсказания заключается в определении параметров  $\{a_k\}$  по речевому сигналу. При этом предполагается, что полученные параметры являются параметрами системной функции  $H(z)$  в модели речеобразования. Вследствие изменения свойств речевого сигнала во времени коэффициенты предсказания должны оцениваться на коротких сегментах речи – кадрах.

В качестве критерия, по которому производится оптимизация синтеза фильтра  $A(z)$ , удобно взять минимум суммы квадратов погрешностей линейного предсказания на сегменте (кадре) речевого сигнала. Основные причины для выбора такого критерия следующие: получающиеся уравнения линейные, относительно просто решаются и дают хорошие результаты.

Пусть  $[n_0, n_1]$  – некоторый интервал. Сумма квадратов погрешностей линейного предсказания определяется следующим образом:

$$E = \sum_{n=n_0}^{n_1} e^2(n).$$

Параметры  $a_k$  можно получить, минимизируя  $E$ . Подставим (1.2) в выражение для  $E$  и приравняем нулю производные  $\partial E / \partial a_k, k = 1, 2, \dots, p$ . Получаем

$$\begin{aligned} E &= \sum_{n=n_0}^{n_1} (s(n) - a_1 s(n-1) - \dots - a_p s(n-p))^2 = \\ &= \sum_{n=n_0}^{n_1} s^2(n) - 2 \sum_{i=1}^p a_i \sum_{n=n_0}^{n_1} s(n)s(n-i) + \sum_{i=1}^p \sum_{j=1}^p a_i a_j \sum_{n=n_0}^{n_1} s(n-i)s(n-j). \end{aligned} \quad (1.3)$$

Дифференцируем (1.3) по  $a_k, k = 1, 2, \dots, p$ ,

$$\partial E / \partial a_k = \sum_{n=n_0}^{n_1} s(n)s(n-k) - \sum_{i=1}^p a_i \sum_{n=n_0}^{n_1} s(n-k)s(n-i) = 0. \quad (1.4)$$

Заменив  $k$  на  $j$ , получим систему  $p$  линейных уравнений относительно  $p$  неизвестных  $a_1, a_2, \dots, a_p$

$$\sum_{i=1}^p a_i c_{ij} = c_{0j}, \quad j = 1, 2, \dots, p, \quad (1.5, a)$$

где

$$c_{ij} = c_{ji} = \sum_{n=n_0}^{n_1} s(n-i)s(n-j). \quad (1.5, б)$$

Данная система называется системой уравнений Юла–Уокера. Имея решение этой системы, нетрудно оценить и минимальную достижимую погрешность предсказания. Для этого подставим (1.5, б) в (1.3)

$$E = c_{00} - 2 \sum_{i=1}^p a_i c_{0i} + \sum_{i=1}^p a_i \sum_{j=1}^p a_j c_{ij},$$

и, используя (1.5, а), упростим это выражение. В результате получим

$$E = c_{00} - \sum_{i=1}^p a_i c_{0i}.$$

Для определения коэффициентов  $a_k$  из уравнений Юла–Уокера необходимо знать величины  $c_{ij}, i = 0, 1, \dots, p, j = 1, 2, \dots, p$ . Имеется два подхода к вычислению этих величин. Один называется ковариационным методом, а второй – автокорреляционным методом.

### Автокорреляционный метод

Для этого метода полагаем пределы анализа равными  $n_0 = -\infty, n_1 = \infty$ , причем сигнал обнуляется вне интервала анализа, т.е.  $s(n) = 0$  при  $n < 0, n \geq N$ . Такие пределы позволяют упростить выражение для  $c_{ij}, i = 1, 2, \dots, p, j = 0, 1, \dots, p$ .

$$c_{ij} = \sum_{n=0}^{N+p-1} s(n-i)s(n-j) = \sum_{n=0}^{N-1-|i-j|} s(n)s(n+|i-j|).$$

В этом случае  $c_{ij}$  являются функциями величины  $|i-j|$  и с точностью до множителя совпадают с оценками автокорреляционной функции  $\hat{R}(\tau)$  сигнала  $s(n)$ , вычисленными при  $\tau = i-j$

$$\hat{R}(|i-j|) = c_{ij} / N = 1/N \sum_{n=0}^{N-1-|i-j|} s(n)s(n+|i-j|).$$

Разделив уравнения в системе (1.5,а) на  $N$ , мы получим систему уравнений Юла–Уокера для автокорреляционного метода

$$\sum_{i=1}^p a_i \hat{R}(|i-j|) = \hat{R}(j), j = 1, 2, \dots, p. \quad (1.6)$$

В матричном виде система может быть записана как  $\mathbf{a} \times \mathbf{R} = \mathbf{b}$ , где

$$\mathbf{a} = (a_1, a_2, \dots, a_p), \quad \mathbf{b} = (\hat{R}(1), \hat{R}(2), \dots, \hat{R}(p)), \quad \mathbf{R} = \begin{bmatrix} \hat{R}(0) & \hat{R}(1) \dots & \hat{R}(p-1) \\ \hat{R}(1) & \hat{R}(0) \dots & \hat{R}(p-2) \\ \dots & \dots & \dots \\ \hat{R}(p-1) & \hat{R}(p-2) \dots & \hat{R}(0) \end{bmatrix}.$$

Матрица  $\mathbf{R}$  в автокорреляционном методе обладает двумя важными свойствами. Она симметрическая (ее элементы, симметричные относительно главной диагонали, равны) и теплицева (каждая следующая строка получается из предыдущей сдвигом вправо). Структура теплицевой матрицы позволяет решить систему (1.6) особенно просто: для определения решения по алгоритму Левинсона–Дарбина, описание которого приводится ниже, требуется порядка  $p^2$  операций (напомним, что решение произвольной системы  $p$  уравнений с  $p$  неизвестными потребовало бы порядка  $p^3$  операций).

### Ковариационный метод

В этом методе выбирается  $n_0 = 0$ ,  $n_1 = N-1$ , а сигнал  $s(n)$  не ограничивается. При этом для величин  $c_{ij}$ ,  $i = 0, 1, \dots, p$ ,  $j = 1, 2, \dots, p$  мы получим

выражение  $c_{ij} = \sum_{n=0}^{N-1} s(n-i)s(n-j)$ . Изменив индекс суммирования, это

выражение можно представить в виде

$$c_{ij} = \sum_{n=-i}^{N-i-1} s(n)s(n+i-j), \quad i = 1, \dots, p, \quad j = 0, \dots, p. \quad (1.7)$$

Выражение (1.7) похоже на выражение для  $c_{ij}$  для автокорреляционного метода, но имеет другие пределы суммирования. В (1.7) используются значения сигнала вне интервала  $0 \leq n \leq N-1$ . Другими словами для вычисления  $c_{ij}$  в ковариационном методе необходимо знать значения сигнала  $s(-p), s(-p+1), \dots, s(N-1)$ , т.е. сигнал должен быть известен на несколько большем интервале, чем в автокорреляционном методе. Однако как правило,  $p \ll N$  и данное требование не очень существенно. Этот метод приводит не к автокорреляционной, а к взаимной корреляционной функции между двумя

очень сходными, но не одинаковыми сегментами речевого сигнала конечной длительности

$$\hat{R}(i, j) = c_{ij} / N = 1 / N \sum_{n=0}^{N-1} s(n-i)s(n-j).$$

Нетрудно видеть, что  $\hat{R}(i, j) = \hat{R}(j, i)$ , однако  $\hat{R}(i, j)$  не является функцией от  $i - j$ , как это было в автокорреляционном методе. Разделив все уравнения в системе (1.4) на  $N$ , мы получим систему уравнений Юла–Уокера для ковариационного метода

$$\sum_{i=1}^p a_i \hat{R}(i, j) = \hat{R}(0, j), j = 1, 2, \dots, p. \quad (1.8)$$

В матричном виде система уравнений (1.8) имеет вид  $\mathbf{a} \times \mathbf{P} = \mathbf{c}$ , где

$$\mathbf{a} = (a_1, a_2, \dots, a_p), \quad \mathbf{c} = (\hat{R}(0,1), \hat{R}(0,2), \dots, \hat{R}(0,p)), \quad \mathbf{P} = \begin{bmatrix} \hat{R}(1,1) & \hat{R}(1,2) & \dots & \hat{R}(1,p) \\ \hat{R}(2,1) & \hat{R}(2,2) & \dots & \hat{R}(2,p) \\ \dots & \dots & \dots & \dots \\ \hat{R}(p,1) & \hat{R}(p,2) & \dots & \hat{R}(p,p) \end{bmatrix}.$$

В отличие от матрицы  $\mathbf{R}$  в автокорреляционном методе матрица  $\mathbf{P}$  будет симметрической, но не теплицевой. Решение такой системы в общем виде требует  $p^3$  операций.

### Алгоритмы решения уравнений линейного предсказания

Рассмотрим алгоритм решения уравнений линейного предсказания в случае автокорреляционного метода. В этом случае система уравнений Юла–Уокера имеет вид (1.6) и матрица коэффициентов  $\mathbf{R}$  является теплицевой и симметрической. Это позволяет найти решение (1.6) за  $p^2$  операций с помощью следующего алгоритма.

#### Алгоритм Левинсона–Дарбина

Данный алгоритм был предложен Левинсоном в 1948 г. и усовершенствован Дарбиным в 1960 г.. Особенностью алгоритма является его итеративный характер. В нем последовательно решается система вида (1.6) порядка  $l = 1, l = 2, \dots, l = p$ , причем решение системы порядка  $l$  выражается через решение системы порядка  $l - 1$ . Решение системы порядка  $l$  будем обозначать через  $\mathbf{a}^{(l)} = (a_1^{(l)}, a_2^{(l)}, \dots, a_l^{(l)})$ . На каждом шаге алгоритма вычисляется также ошибка предсказания  $E_l$  для решения системы  $l$ -го порядка и вспомогательный коэффициент  $k_l$ . Ниже приводится формальное описание алгоритма.

**Начальные условия**

$$l = 0, E_0 = R(0), a^{(0)} = 0.$$

**Итеративная процедура**

При  $l = 1, \dots, p$  вычисляются

$$k_l = \left( \sum_{i=1}^{l-1} a_i^{(l-1)} R(l-i) - R(l) \right) / E_{l-1}, \quad (1.9,а)$$

$$a_l^{(l)} = -k_l,$$

$$a_j^{(l)} = a_j^{(l-1)} + k_l a_{l-j}^{(l-1)}, \quad 1 \leq j \leq l-1,$$

$$E_l = E_{l-1} (1 - k_l^2). \quad (1.9,б)$$

На последнем шаге алгоритма при  $l = p$  мы должны получить искомое решение  $\mathbf{a} = (a_1, a_2, \dots, a_p) = \mathbf{a}^{(p)}, E = E_p$ .

**Пример.** Проследим работу алгоритма Левинсона–Дарбина на первых шагах.

**Шаг 1**

$$k_1 = -R(1)/R(0), \quad a_1^{(1)} = R(1)/R(0), \quad E_1 = (R^2(0) - R^{(2)}(1))/R(0).$$

**Шаг 2**

$$k_2 = (R^2(1) - R(0)R(2))/(R^2(0) - R^{(2)}(1)),$$

$$a_1^{(2)} = (R(1)R(0) - R(1)R(2))/(R^2(0) - R^2(1))$$

$$E_2 = (R(0) - R(2))(R^2(0) + R(2)R(0) - 2R^2(1))/(R^2(0) - R^2(1)).$$

**Устойчивость авторегрессионной модели**

При синтезе случайного процесса с помощью полученного фильтра необходимо предварительно рассмотреть вопрос об устойчивости. На практике отсутствие устойчивости проявляется в том, что моделирование с построенным фильтром не даст нужного результата, т.е. малым сигналам возбуждения на входе фильтра могут соответствовать большие выходные сигналы. Ниже мы рассмотрим некоторые вопросы, связанные с устойчивостью авторегрессионного фильтра и порождаемого им процесса.

Последовательность  $\{x(n)\} = x(1), x(2), \dots, x(n), \dots$  называется устойчивой, если  $X(z) = \sum_{n=1}^{\infty} x(n)z^n$  абсолютно сходится внутри единичного круга, т.е. при  $|z| < 1$ . Фильтр  $H(z)$  называется устойчивым, если все полюсы  $H(z)$  лежат внутри единичного круга. Из этих определений следует, что если на вход фильтра поступает устойчивая последовательность и фильтр устойчив, то на выходе будет также устойчивая последовательность.

Действительно, мы имеем  $|X(z)| = |Y(z)||H(z)| < \infty$  при всех  $z < 1$  тогда и только тогда, когда  $|Y(z)| < \infty$  и  $|H(z)| < \infty$  при  $z < 1$ . Следовательно, для проверки устойчивости фильтра с передаточной функцией  $H(z) = 1/A(z)$  нужно вычислить все корни полинома  $A(z) = 1 - a_1 z^{-1} - \dots - a_p z^{-p}$  и убедиться в том, что они удовлетворяют условию  $|z_i| < 1, 1 \leq i \leq p$  (для того чтобы привести  $A(z)$  к полиномиальному виду, можно умножить  $A(z)$  на  $z^p$ ).

Процедура вычисления всех комплексных корней многочлена достаточно трудоемка и на практике применяется редко. Однако если фильтр синтезируется по алгоритму Левинсона–Дарбина, то условием устойчивости будет выполнение на каждом шаге неравенства  $|k_l| < 1$ .

### ЗАДАНИЯ

1. Получить у преподавателя файл с речевым сигналом, записанным в формате wav.
2. Написать программу построения фильтра по методу Левинсона–Дарбина для кадров речевого сигнала размером 180–240 отсчетов.
3. Вычислить "идеальное" возбуждение для системы рис.1.1 путем фильтрации речевого сигнала обратным фильтром с передаточной функцией  $A(z)$ .
4. Выполнить равномерное скалярное квантование коэффициентов полученного фильтра.
5. Получить синтезированный речевой сигнал путем пропускания "идеального" возбуждения через фильтр с квантованными коэффициентами. Построить график зависимости ошибки аппроксимации речевого сигнала от числа бит, затрачиваемых на хранение коэффициентов фильтра. Определить число бит, при котором достигается высокое качество синтезированного сигнала.
6. Выполнить равномерное скалярное квантование сигнала возбуждения.
7. Получить синтезированный речевой сигнал путем пропускания квантованного возбуждения через фильтр с квантованными коэффициентами. Построить график зависимости ошибки аппроксимации речевого сигнала от суммарного числа бит, затрачиваемых на хранение или передачу речевого сигнала. Определить скорость передачи в бит/с, при которой достигается высокое качество аппроксимации речевого сигнала.