Лекция 6
Кодеры с “анализом-через-синтез” на основе линейного предсказания (LPAS кодеры)

В настоящее время большинство речевых кодеров, функционирующих в диапазоне битовых скоростей от 4.8 до 16 Кбит/с, представляют собой кодеры, основанные на модели речеобразования, и используют метод линейного предсказания с анализом-через-синтез. В модели речеобразования для того, чтобы синтезировать речевой сигнал на вход адаптивного линейного предсказывающего фильтра поступает соответствующий сигнал возбуждения. Параметры фильтра и сигнала возбуждения оцениваются и обновляются через регулярные промежутки времени, называемые кадрами. Сжатый речевой файл содержит параметры модели, оцененные для всех кадров.

На Рис. 6.1 показана модель речеобразования. Фильтр линейного предсказания моделирует изменения, которые происходят в речевом тракте, сформированном губами, языком и зубами при произнесении того или иного звука. Грубо говоря, каждый звук соответствует некоторому множеству коэффициентов фильтра. Часто фильтр представляется полюсами своей частотной характеристики, называемыми *формантными частотами* или *формантами*. Возбуждение фильтра зависит от типа произносимого звука: звонкий, глухой, гласный, шипящий или носовой. Звонкие звуки порождаются колебаниями голосовых связок. В этом случае сигнал возбуждения представляет собой квази-периодическую импульсную последовательность. Глухие звуки порождаются шумоподобными входными сигналами. Простейшая модель, показанная на Рис.6.1, использует два типа входного сигнала возбуждения (периодическую импульсную последовательность и квази-случайный шум), переключаемые на каждом кадре. Период колебания голосовых связок называется *периодом основного тона* (pitch). Период основного тона оценивается по речевому сигналу и определяет период импульсов в импульсной последовательности. Мы будем рассматривать кодеры с анализом-через-синтез на основе линейного предсказания, использующие адаптацию по истинным (forward adaptive) и восстановленным (backward adaptive) отсчетам.

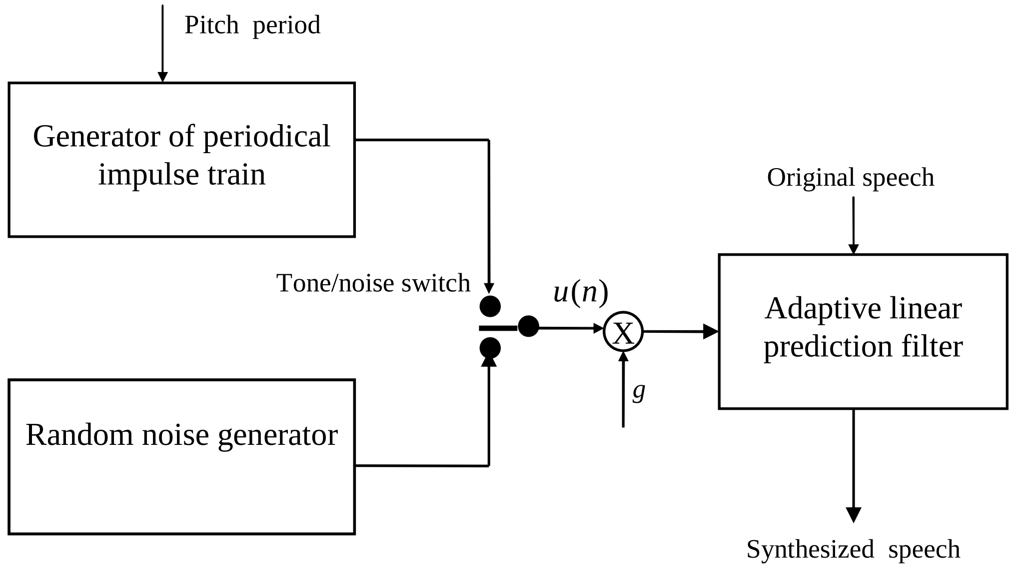

Рис. 6.1 Модель речеобразования

Прежде чем переходить к рассмотрению кодеров, покажем, что задача нахождения оптимальных коэффициентов линейного предсказания может интерпретироваться как задача построения цифрового фильтра.

Как было показано в лекции 5, для нахождения оптимальных коэффициентов линейного предсказания необходимо решить уравнения Юла-Уокера :

$$\sum_{i=1}^{m} a_i c_{ij} = c_{0j}\ ,\ \ j = 1,...,m\ . \qquad (6.1)$$

Предположим, что коэффициенты a_i найдены. Тогда ошибка предсказания определяется как

$$e(n) = x(n) - \sum_{i=1}^{m} a_i x(n-i)\ . \qquad (6.2)$$

Нетрудно видеть, что уравнение (6.2) описывает КИХ фильтр с системной функцией

$$A(z) = \frac{E(z)}{X(z)} = 1 - \sum_{i=1}^{m} a_i z^{-i}\ .$$

С другой стороны, при известной ошибке предсказания $e(n)$ синтезированный речевой сигнал представляет собой выход БИХ фильтра, описываемого уравнением

$$x_s(n) = e(n) + \sum_{i=1}^{m} a_i x(n-i)\ .$$

Если $e(n)$ – это неквантованная ошибка предсказания, а a_i – неквантованные коэффициенты предсказания, то $x_s(n) = x(n)$. Таким образом, речевой сигнал формируется на выходе цифрового фильтра с системной функцией

$$H(z) = \frac{1}{A(z)} = \frac{X(z)}{E(z)} = \frac{1}{1 - \sum_{i=1}^{m} a_i z^{-i}}.$$

В речевых кодерах рассматриваемого типа на каждом кадре для нахождения оптимальных коэффициентов предсказания необходимо решать систему уравнений Юла-Уокера. Так как стандартный размер кадра составляет 180-240 отсчетов, то есть при частоте дискретизации 8 кГц система уравнений (6.1) решается каждые 25-30 мс, то сложность решения этой системы существенно влияет на быстродействие кодера в целом. Для определения коэффициентов a_i из уравнений (6.1) необходимо знать величины c_{ij}, $i = 0,1,...,m$, $j = 1,...,m$. Оказывается, что сложность решения системы (6.1) зависит от метода определения этих величин.

АВТОКОРРЕЛЯЦИОННЫЙ МЕТОД НАХОЖДЕНИЯ КОЭФФИЦИЕНТОВ c_{ij}

Напомним, что $c_{ij} = \sum_{n=n_0}^{n_1} x(n-i)x(n-j)$. В этом методе полагаем, что сигнал обнуляется вне окна анализа, равного по длине кадру речевого сигнала, т.е. $x(n) = 0$ при $n < 0$, $n \geq N$, где N – размер кадра. Такие пределы позволяют упростить выражение для c_{ij}, $i = 0,1,...,m$, $j = 1,...,m$

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

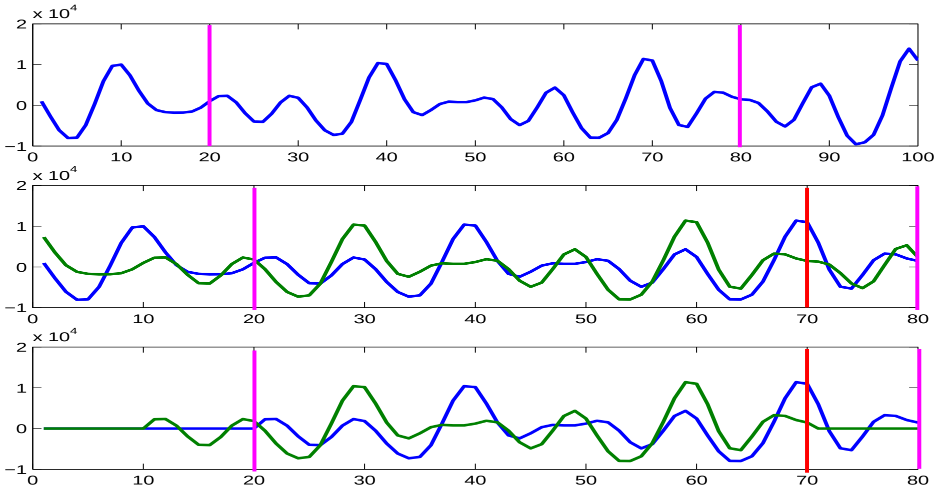

Рис. 6.2 Автокорреляционный метод

На Рис. 6.2 размер кадра $N = 60$. Рассматривается речевой сигнал и его сдвиг на 10 отсчетов. Нетрудно видеть, что, благодаря обнулению сигнала вне кадра, все отсчеты с номерами меньше 20 (0-й отсчет кадра) и больше 70 ($(N-1-10)$-й отсчет кадра) умножаются на ноль.

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

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

Разделив уравнения в системе (6.1) на N, мы получим систему уравнений Юла-Уокера для автокорреляционного метода

$$\sum_{i=1}^{m} a_i R(|i-j|) = R(j),\ \ j = 1,...,m. \qquad (6.3)$$

В матричном виде система может быть записана как

$$\mathbf{aR} = \mathbf{b},$$

где $\mathbf{a} = (a_1, ..., a_m)$, $\mathbf{b} = (R(1), R(2), ..., R(m))$,

$$\mathbf{R} = \begin{pmatrix} R(0) & R(1) & ... & R(m-1) \\ R(1) & R(0) & ... & R(m-2) \\ ... & ... & ... & ... \\ R(m-1) & R(m-2) & ... & R(0) \end{pmatrix}.$$

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 (6.3) особенно просто: для определения решения по алгоритму Левинсона-Дарбина, описание которого приводится ниже, требуется порядка m^2 операций. Напомним, что решение произвольной системы m уравнений с m неизвестными потребовало бы порядка m^3 операций.

АЛГОРИТМ ЛЕВИНСОНА-ДАРБИНА РЕШЕНИЯ УРАВНЕНИЙ ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ

Рассмотрим алгоритм решения уравнений линейного предсказания в случае автокорреляционного метода. Этот метод учитывает, что матрица коэффициентов $\mathbf{R}$ является теплицевой и симметрической.

Данный алгоритм был предложен Левинсоном в 1948 году и усовершенствован Дарбиным в 1960 году. Особенностью алгоритма является его итеративный характер. В нем последовательно решается система уравнений (6.3) порядка $l = 1$, $l = 2, ...$, $l = m$, причем решение системы порядка l выражается через решение системы порядка $(l-1)$. Решение системы порядка l будем обозначать через $\mathbf{a}^{(l)} = (a_1^{(l)}, a_2^{(l)}, ..., a_l^{(l)})$. 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 $l = 1, ..., m$ вычисляются

$$a_l^l = \frac{R(l) - \sum_{i=1}^{l-1} a_i^{(l-1)} R(l-i)}{E_{l-1}},$$

$$a_j^{(l)} = a_j^{(l-1)} - a_l^l a_{l-j}^{(l-1)}, \; 1 \le j \le l-1,$$

$$E_l = E_{l-1}(1 - (a_l^l)^2).$$

На последнем шаге алгоритма при $l = m$ мы получаем искомое решение $\mathbf{a} = (a_1, a_2, ..., a_m) = \mathbf{a}^{(m)}$, $E = E_m$.

Заметим, что число коэффициентов, вычисляемых алгоритмом, при $l = 1, 2, ..., m$ образует арифметическую прогрессию: $1, 2, ..., m$. Сумма этой прогрессии равна $(m+1)m/2$. Отсюда следует, что сложность алгоритма Левинсона-Дарбина имеет порядок m^2.

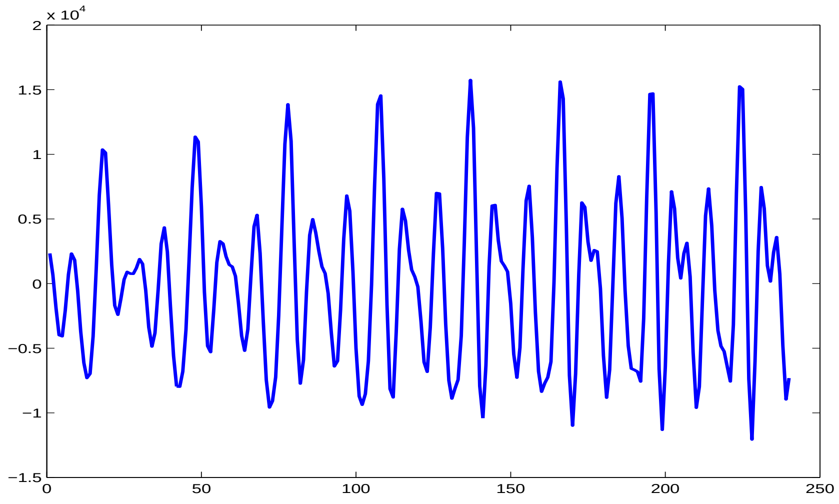

Рис. 6.3 Кадр речевого сигнала.

На Рис.6.3 показан кадр речевого сигнала, содержащий 240 отсчетов. Нормированные коэффициенты $R(0),\ldots,R(10)$ в системе уравнений Юла-Уокера равны:

1.0, 0.75, 0.17, –0.38, –0.65, –0.59, –0.35, –0.08, 0.17, 0.39, 0.52, $R(0) = 3.79 \times 10^7$.

Коэффициенты фильтра $a_1, a_2, \ldots, a_{10}$ равны

1.989, –1.734, 0.412, 0.096, 0.128, –0.084, –0.378, 0.303, 0.131, –0.166, соответственно.

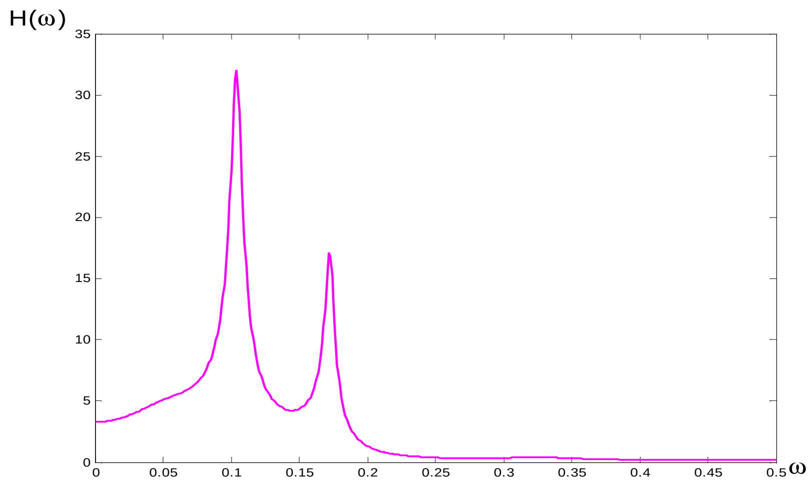

Рис. 6.4 АЧХ фильтра в модели речеобразования

КОДИРОВАНИЕ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ПАРАМЕТРОВ

В кодерах с “анализом-через-синтез”, использующих линейное предсказание, сжатый файл содержит квантованные параметры модели речеобразования. Таким образом, после нахождения оптимальных коэффициентов предсказания для данного кадра, их необходимо проквантовать. Экспериментально установлено, что каждый из коэффициентов меняется в достаточно широком диапазоне и, что коэффициенты существенно меняются от кадра к кадру, даже, если речевой сигнал на этих кадрах меняется незначительно. По этой причине часто от описания фильтра линейного предсказания во временной области переходят к его описанию в частотной области, основанному на вычислении так называемых *линейных спектральных параметров (ЛСП).* Хотя эти параметры однозначно вычисляются по коэффициентам фильтра, квантование ЛСП оказывается предпочтительнее квантования коэффициентов. ЛСП фильтра порядка m представляют собой упорядоченный набор из m чисел, принимающих значения из конечного интервала $[0, f_s/2]$, где f_s – частота дискретизации речевого сигнала. Диапазон изменения значений каждого из ЛСП относительно невелик, кроме того, на вокализованных участках речи ЛСП меняются относительно медленно во времени по сравнению с коэффициентами предсказывающего фильтра.

Предположим, что нам известны коэффициенты предсказывающего фильтра $\mathbf{a} = (a_1, a_2, ..., a_m)$, тогда мы можем записать системную функцию синтезирующего фильтра в виде

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{m} a_i z^{-i}}.$$

Так как коэффициенты фильтра – вещественные числа, то $A(z)$ имеет попарно комплексно сопряженные корни $z_{\pm i} = r_i e^{\pm j\omega_i}$, $i = 1,...,m/2$.

Тогда амплитудно-частотная характеристика (АЧХ) фильтра имеет вид

$$\left|H(e^{j\omega})\right| = \frac{1}{\left|A(e^{j\omega})\right|} = \frac{1}{\left|1 - \sum_{i=1}^{m} a_i e^{-j\omega}\right|} = \frac{1}{\left|\prod_{k=1}^{m/2} (z - r_k e^{j\omega_k})(z - r_k e^{-j\omega_k})\right|}.$$

Частоты $\omega_1,...,\omega_{m/2}$, соответствующие нулям полинома $A(z)$ (резонансам АЧХ), называют *формантными частотами.* На Рис. 6.4 показана АЧХ синтезирующего фильтра, построенного по кадру речевого сигнала Рис. 6.3. На Рис. 6.4 хорошо видны резонансные пики, соответствующие первым двум формантным частотам, остальные резонансные пики, соответствующие еще трем формантным частотам выражены не резко. ЛСП – это частотные параметры речевого сигнала, определенным образом связанные с его формантными частотами и получаемые по описанному ниже алгоритму.

Полином $A(z)$ степени m имеет в точности m корней ($z_{\pm i} = r_i e^{\pm j\omega_i}$, $i = 1,...,m/2$). Если фильтр устойчив, то его корни могут быть представлены в виде точек, лежащих внутри единичной окружности на плоскости комплексных чисел. Задача нахождения комплексных корней полинома высокого порядка имеет высокую вычислительную сложность. Цель описываемого далее алгоритма свести задачу нахождения комплексных корней полинома m-го порядка к задаче нахождения вещественных корней двух вспомогательных полиномов порядка $m/2$.

На первом шаге алгоритма по полиному $A(z)$ строятся два вспомогательных полинома степени $m+1$ – суммарный и разностный, каждый из которых имеет корни, лежащие на единичной окружности. Каждый из полученных полиномов имеет тривиальный вещественный корень, что позволяет понизить порядок полиномов до m. Далее, так как корни вспомогательных полиномов $e^{\pm j\omega_n}$, $n = 1,...,m/2$ и $e^{\pm j\omega_k}$, $k = 1,...,m/2$ лежат на единичной окружности, т.е. имеют единичный модуль, они могут быть заданы величинами фаз. Комплексно сопряженным корням соответствуют противоположные фазы. Это наблюдение позволяет простыми преобразованиями свести каждое из уравнений степени m к уравнению степени $m/2$ относительно косинусов фаз корней. Вычислив арккосинусы, мы получаем ровно m чисел φ_i, $i=1,...,m$ на отрезке $[0,\pi]$. Для приведения значений к отрезку $[0, f_s]$ выполняем нормировку $\phi_i = \varphi_i f_s/(2\pi)$. Полученные значения называют *линейными спектральными параметрами*. Ниже приведено детальное описание алгоритма вычисления ЛСП по коэффициентам фильтра, а затем восстановления коэффициентов фильтра по известным ЛСП.

Шаг 1. Построение суммарного и разностного полиномов $P(z)$ и $Q(z)$ на основе исходного полинома $A(z)$. Представим $A(z)$ в виде

$$A(z) = A(1)z^m + A(2)z^{m-1} + ... + A(m)z + A(m+1),$$

где $A(1)=1, A(2)=-a_1,..., A(m+1)=-a_m$. Построим суммарный полином

$$P(z) = P(1)z^{m+1} + P(2)z^m + ... + P(m+1)z + P(m+2)$$

степени $m+1$ по правилу:

$$P(1)=1, P(m+2)=1, P(k) = A(k) + A(m+3-k)$$

для $k=2,...,m/2+1$,

$$P(k) = P(m+3-k)$$

для $k = m/2+2,...,m+1$. Например, для фильтра порядка $m=2$ мы получаем

$$A(z) = z^2 + A(2)z + A(3),$$

и

$$P(z) = z^3 + (A(2)+A(3))z^2 + (A(2)+A(3))z + 1.$$

Построим разностный полином

$$Q(z) = Q(1)z^{m+1} + Q(2)z^m + ... + Q(m+1)z + Q(m+2)$$

степени $m+1$ по правилу:

$$Q(1)=1, Q(m+2)=-1, Q(k) = A(k) - A(m+3-k) \text{ для } k=2,...,m/2+1,$$

$$Q(k) = -Q(m+3-k) \text{ для } k = m/2+2,...,m+1.$$

Например, для фильтра порядка $m=2$ получаем

$$Q(z) = z^3 + (A(2)-A(3))z^2 - (A(2)-A(3))z - 1.$$

Нетрудно видеть, что $A(z) = (P(z)+Q(z))/2z$.

Шаг 2. Понижение порядка полиномов на 1. Строим суммарный полином $PL(z)$ порядка m из полинома $P(z)$ по правилу:

$$PL(1)=1,\ PL(k) = P(k) - PL(k-1)$$

для $k=2,...,m/2+1$,

$$PL(k) = PL(m+2-k)$$

для $k = m/2+2,...,m+1$.

Отметим, что при выполнении условий устойчивости полученный полином имеет только комплексные корни и эти корни лежат на единичной окружности.

Например, для фильтра порядка $m=2$

$$PL(z) = z^2 + (A(2)+A(3)-1)z + 1,$$

что эквивалентно делению $P(z)$ на $(z+1)$.

Построим разностный полином $QL(z)$ степени m по правилу:

$$QL(1)=1,\ QL(k) = Q(k) + QL(k-1)$$

для $k=2,...,m/2+1$,

$$QL(k) = QL(m+2-k),$$

для $k = m/2+2,...,m+1$. Можно показать, что в случае устойчивого фильтра корни полинома $QL(z)$ – комплексные величины и лежат на единичной окружности.
Для фильтра порядка $m = 2$

$$QL(z) = z^2 + (A(2) - A(3) + 1)z + 1,$$

что эквивалентно делению $Q(z)$ на $(z-1)$.

Шаг 3. Понижение порядка до $m/2$. Учитывая, что корни полиномов $PL(z)$ и $QL(z)$ лежат на единичной окружности , т.е. имеют вид $z = e^{\pm j\omega_i}$, где $i = 1,...,m/2$, мы можем понизить порядок уравнения, которое нужно решить для нахождения корней $QL(z) = 0$ и $PL(z) = 0$. Строим полиномы $P^*(z)$ и $Q^*(z)$ степени $m/2$ по формулам:

$$P^*(m/2+1) = 2^{-m/2}(PL(m/2+1) - 2PL(m/2-1) + 2PL(m/2-3) - ...),$$
$$P^*(m/2) = 2^{-(m/2-1)}(PL(m/2) - 3PL(m/2-2) + 5PL(m/2-4) - ...),$$
$$P^*(m/2-1) = 2^{-(m/2-2)}(PL(m/2-1) - 4PL(m/2-3) + 9PL(m/2-5) - ...),$$
....
$$P^*(2) = PL(2)/2,$$
$$P^*(1) = 1.$$

Формальное правило, по которому записаны коэффициенты следующее. В формуле для $P^*(m/2+1)$ в скобках за коэффициентом 1 следует последовательность из чередующихся коэффициентов –2 и +2. Во всех следующих формулах первый коэффициент равен 1. Знаки следующих коэффициентов чередуются. Коэффициент при i-м члене каждой последовательности, $i > 1$ равен по абсолютной величине сумме абсолютных величин $(i-1)$-го коэффициента данной формулы и i-го коэффициента предыдущей формулы. Например, в формуле для $P^*(m/2-1)$ имеем 4=3+1, 9=4+5, далее последует 16=7+9 и т. д.

Коэффициенты полинома $Q^*(z)$ могут быть получены аналогично из коэффициентов полинома $QL(z)$. В частности, для фильтра порядка $m = 2$ получаем

$$P^*(z) = z + (A(2) + A(3) - 1)/2 \ , \ Q^*(z) = z + (A(2) - A(3) + 1)/2 .$$

Шаг 4. Решаем уравнения $Q^*(z) = 0$ и $P^*(z) = 0$. Эти уравнения имеют $m/2$ вещественных корней вида $z_i = \cos\omega_i$, где $i = 1,2,...,m/2$. Для фильтра порядка $m = 2$ получаем

$$zp = -(A(3) + A(2) - 1)/2 , \ \ zq = -(A(2) - A(3) + 1)/2 .$$

Для решения уравнений произвольного порядка $Q^*(z) = 0$ и $P^*(z) = 0$ можно использовать любой метод численного решения алгебраических уравнений. На практике часто используют следующий метод. Поиск корней полинома производится в два этапа. Сначала на интервале –1, 1 ищут приближенные значения корней $root(j), j = 1,2,...,m/2$. Для этого интервал делится на n равных интервалов, где величина n зависит от порядка фильтра. Например, для $m = 10$ выбирают $n = 128$. Наличие корня в интервале обнаруживается по изменению знака полинома. На втором этапе производится уточнение значения корня. Уточненное значение корня $exroot(j), j = 1,2,...,m/2$ вычисляется по формуле

$$exroot(j) = root(j) + 2F(i-1)/(n(F(i-1) - F(i))\,,$$

где i – номер интервала, на котором располагается корень с номером j;

$root(j) = -1 + 2(i-1)/n$; $F(i)$ – значение полинома в конце i -го интервала.

Шаг 5. Обозначим через $zp_i, i = 1,2,...,m/2$ корни полинома $P^*(z)$ и через $zq_i, i = 1,2,...,m/2$ корни полинома $Q^*(z)$. Находим ЛСП по формулам

$$\omega p_i = \arccos(zp_i),\ i = 1,...m/2,$$
$$\omega q_i = \arccos(zq_i),\ i = 1,...m/2.$$

Далее выполняется сортировка значений по возрастанию и их нормировка умножением на частоту дискретизации и делением на 2π.

Полученные ЛСП могут быть проквантованы скалярно или векторно и закодированы вместе с другими параметрами речевого сигнала. На приемной стороне решается задача восстановления коэффициентов фильтра по квантованным значениям ЛСП по следующему алгоритму.

Шаг 1. Квантованные ЛСП разбиваются на подмножества, одно из которых (значения с четными номерами) соответствует суммарному полиному, другое – разностному. Значения умножаются на 2π и делятся на частоту дискретизации. Вычисляется косинус каждого значения. Тем самым найдены корни полиномов $P^*(z)$ и $Q^*(z)$.

Шаг 2. Используя обобщение теоремы Виета, можно выразить все коэффициенты полиномов $P^*(z)$ и $Q^*(z)$ через их корни. Напомним, что для произвольного полинома порядка n

$$a_n x^n + a_{n-1}x^{n-1} + a_1 x + a_0,$$
$$a_{n-i} = (-1)^i s_i a_n,$$

где

$$s_0 = 1,$$
$$s_1 = r_1 + r_2 + r_3 + r_4 + \dots,$$
$$s_2 = r_1r_2 + r_1r_3 + r_1r_4 + r_2r_3 + \dots,$$
$$s_3 = r_1r_2r_3 + r_1r_2r_4 + r_2r_3r_4 + \dots,$$
$$\dots$$
$$s_n = r_1r_2r_3.....r_n.$$

Для фильтра порядка $m = 2$ получаем $P^*(2) = -zp, Q^*(2) = -zq, P^*(1) = 1, Q^*(1) = 1$.

Шаг 3. Пользуясь уже приведенными формулами, связывающими коэффициенты полиномов $P^*(z)$ и $PL(z)$ ($Q^*(z), QL(z)$), рекуррентно восстанавливаем коэффициенты полиномов $PL(z), QL(z)$. Для порядка $m = 2$ получаем:

$$PL(1) = 1, PL(2) = -2zp, PL(3) = 1\ ,$$
$$QL(1) = 1, QL(2) = -2zq, QL(3) = 1.$$

Шаг 4. Восстанавливаем суммарный и разностный полиномы $P(z), Q(z)$ порядка $m+1$ по правилам:

$$P(1) = PL(1), P(j) = PL(j) + PL(j-1)$$

для $j = 2,...,m/2+1$,

$$P(j) = P(m+3-j)$$

для $j = m/2+2,...,m+2$,

$$Q(1) = 1, Q(j) = QL(j) - QL(j-1)$$

для $j = 2,\ldots,m/2+1$,

$$Q(j) = -Q(m+3-j)$$

для $j = m/2+2,\ldots,m+2$. Для фильтра порядка $m=2$ получаем

$$P(1)=1, P(2)=-2zp+1=A(2)+A(3), P(3)=P(2), P(4)=1,$$

$$Q(1)=1, Q(2)=-2zq-1=A(2)-A(3), Q(3)=A(3)-A(2), Q(4)=-1.$$

Таким образом, восстановлены полиномы

$$P(z)=z^3+(A(2)+A(3))z^2+(A(2)+A(3))z+1\ ,$$

$$Q(z)=z^3+(A(2)-A(3))z^2+(A(3)-A(2))z-1.$$

Шаг 5. Восстанавливаем полином $A(z)$:

$$A(z)=(P(z)+Q(z))/2z\ .$$

Нетрудно видеть, что при $m=2$ эта формула приводит к правильному результату.