УДК 534.782.001:621.39(045)

**Нечипорук О.П., Одарченко Р.С.,
Потапов В.Г., Тараненко А.Г., Ткалич О.П.**

КОДИРОВАНИЕ РЕЧИ С КОМБИНИРОВАННЫМ ПРЕДСКАЗАНИЕМ

В работе [1] рассмотрен метод кодирования речи, применяемый в системе сотовой связи с кодовым разделением каналов CDMA (Code Division Multiple Access). Альтернативой данному методу является метод кодирования речи, применяемый в системе сотовой связи стандарта GSM (Global System for Mobile Communications).

Кодирование речи как преобразование сигналов удобно описывать на основе обобщенной модели речевого процесса, показанной на рис. 1.

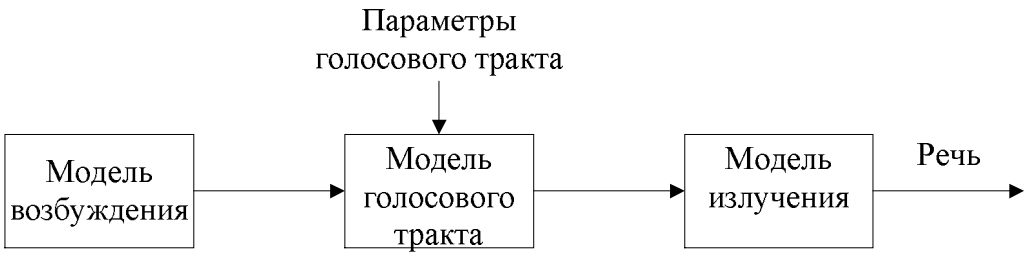

Рис. 1. Обобщенная модель речевого процесса

Математическое моделирование такого процесса может выполняться различными методами. В системах цифровой связи для решения этой задачи широко применяется метод линейного предсказания LPC (Linear Predictive Coding).

Метод основан на математическом представлении дискретного отсчета речевого сигнала в виде выражения

$$U_p(t_k) = GU_b(t_k) + \sum_{i=1}^{m} a_i U_p(t_{k-1}), \quad (1)$$

где t_k - скользящий момент времени в интервале наблюдения сигнала,

G – коэффициент усиления,

$U_b(t_k)$ - сигнал возбуждения голосового тракта в момент t_k,

a_i - коэффициенты, характеризующие голосовой тракт,

m – порядок модели, т.е. количество предыдущих отсчетов речевого сигнала. Очевидно, что отсчет $U_p(t_k)$ определяется отсчетом сигнала возбуждения и взвешенной суммой m предыдущих отсчетов речи.

Уравнение (1) поясняется рис.2.

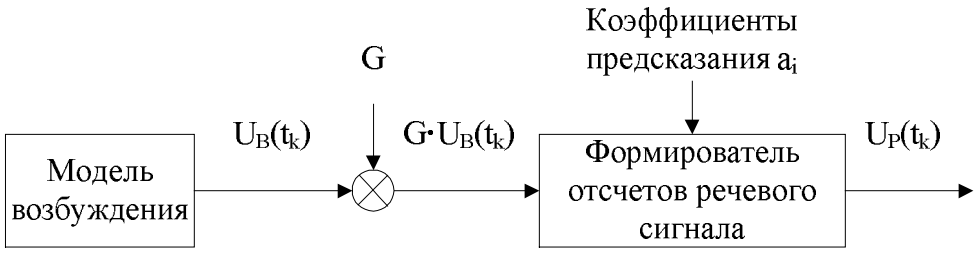

Рис. 2. Линейное предсказание речи

Предыдущие отсчеты речевого сигнала до момента t_k могут быть изменены непосредственно, а сигнал возбуждения первоначально неизвестен. Поэтому предсказываемый (синтезируемый) отсчеты речевого сигнала записывают в виде

$$U_p'(t_k) = \sum_{i=1}^{m} b_i U_p(t_{k-i}), \quad (2)$$

где b_i - оптимальные оценки коэффициентов a_i.

Задача LPC – анализа состоит в вычислении таких коэффициентов b_i, которые при заданном m обеспечивают минимальную среднеквадратическую ошибку предсказания, т.е. минимальную погрешность представления реального речевого сигнала синтезированным сигналом:

$$E(t_k) = \bar{e}^2(t_k) = M\{U_p(t_k) - U_p'(t_k)\}^2 , \quad (3)$$

где M – операция математического ожидания.

Критерий оптимальности коэффициентов b_i состоит в минимализации ошибки (3), что достигается при выполнении условий

$$\frac{dE(t_k)}{db_i} = 0, i = \overline{1, m}. \quad (4)$$

В полученной системе m дифференциальных уравнений коэффициентами при неизменяемых $b_1, \dots, b_m$ являются коэффициенты корреляции дискретных отсчетов речевого сигнала, при этом в практическом кодировании m=8…10.

Данная система решается методами корреляционного анализа; а результатом решения является итерационная процедура вычисления коэффициентов b_i [2].

Рассмотрим рис.3, поясняющий принцип LPC-анализа и LPC-синтеза.

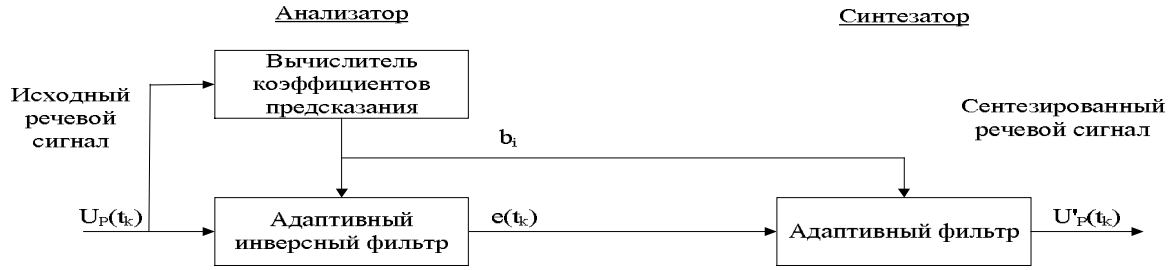

Рис. 3. Анализ и синтез речевого сигнала

По коэффициентам b_i, вычисленным в анализаторе, настраивается адаптивный фильтр в синтезаторе. Этот фильтр имеет передаточную характеристику, отображающую свойства моделируемого голосового тракта, и синтезирует отсчеты речевого сигнала из своего входного сигнала.

Покажем, что в качестве такого входного сигнала можно использовать сигнал ошибки предсказания (остаток предсказания). По физическому смыслу формула (1) характеризует изменения, которые происходят с возбуждающим воздействием при его прохождении через голосовой тракт. Сигнал возбуждения входит только в первое слагаемое формулы (1), поэтому логично полагать второе слагаемое моделью голосового тракта.

Следовательно, LPC-анализ является максимально точным при равенстве коэффициентов a_i и b_i в формулах (1) и (2).

При этом ошибка предсказания равна:

$$e(t_k) = U_p(t_k) - U_p'(t_k) = G \cdot U_b(t_k) \quad (5)$$

Т.е. остаток предсказания в речевом анализаторе является возбуждающим сигналом для речевого синтезатора.

Для получения этого остатка речевой сигнала в анализаторе необходимо пропустить через адаптивный инверсный фильтр с передаточной характеристикой, обратной характеристике синтезирующего фильтра.

Таким образом, метод LPC позволяет выполнить взаимное преобразование речевого сигнала и сигнала возбуждения.

Рассмотренный метод применяется для обработки случайных стационарных сигналов, однако реальный речевой сигнал является нестационарным.

В этом случае речевой сигнал должен подвергаться кратковременному анализу на интервалах квазистационарности τ.

На практике в анализаторе входной сигнал разделяется на фрагменты длительности $\tau \leq 20 \ldots 30$ мс и коэффициенты b_i пересчитываются для каждого нового фрагмента.

Обобщенная структурная схема речевого кодека показана на рис.4.

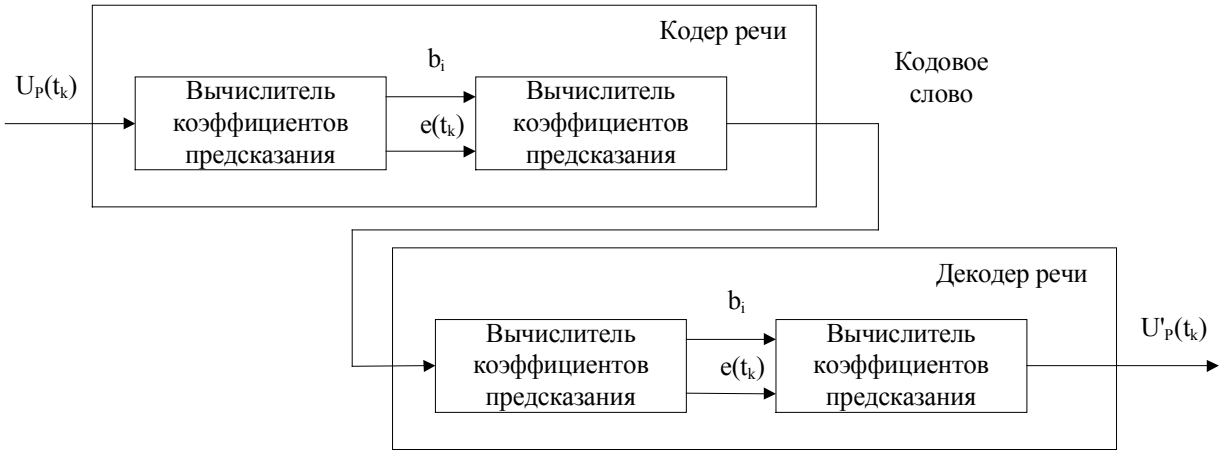

Рис. 4. Обобщенный речевой кодек

Анализатор в составе кодека речи вычисляют коэффициенты для настройки адаптивных фильтров и преобразует речевой сигнал в сигнал возбуждения. Кодер параметров формирует цифровые слова, передаваемые по каналу связи. Принятые слова после декодирования определяют настройку цифрового синтезатора, воспроизводящего речевой сигнал.

Примером речевого кодека, в котором используются преобразования речевого сигнала на основе линейного предсказания, является кодек системы сотовой связи стандарта GSM.

Данный кодек создан на основе двух кодеков. Первый из них реализует способ кодирования RPE-LPC (Regular Pulse Excitation Linear Predictive Coding – кодирование с регулярным импульсным возбуждением и линейным предсказанием). Этот кодек обеспечивает приемлемое качество речи при низкой сложности. Однако на это качество сильно влияют ошибки при передаче информации по каналу связи и акустические помехи, возникающие при синтезировании высокочастотных звуков речи. Второй кодек реализует способ MPE-LTP (Multi-Pulse Excitation Long Term Prediction – кодирование с многоимпульсным возбуждением и долговременным предсказанием). Этот кодек обеспечивает хорошее качество речи при слабом влиянии ошибок, но имеет более высокую сложность. По результатам испытаний первый кодек был дополнен долговременным предсказателем от второго кодека и принят в стандарте GSM под названием RPE-LTP.

Структурная схема кодера приведена на рис. 5.

На вход кодера от аналого-цифрового преобразователя поступает цифровой сигнал в виде двоичных 13-разрядных отсчетов со скоростью передачи 104 кбит/с. Кодирование включает 3 операции: кратковременный анализ, долговременное предсказание и формирование сигнала возбуждения.

В процессе кратковременного анализа входной сигнал разделяется на фрагменты, длительность которых равна интервалу квазистационарности речи 20 мс, при этом в одном интервале содержится 160 цифровых отсчетов. Анализ заключается в линейном предсказании сигнала с использованием модели 8-го порядка. В каждом интервале 20 мс вычисляются коэффициенты предсказания b_k, k=$\overline{1,8}$, по которым настраивается адаптивный фильтр-анализатор. На выходе этого фильтра формируются 160 цифровых отсчетов остатка предсказания $e_ц(t_i)$. Выходными параметрами кратковременного анализа являются коэффициенты b_k.

Для последующих операций текущий интервал 20 мс разделяется на 4 части по 5 мс, т.е. 160 отсчетов остатка разделяются на 4 группы по 40 отсчетов. Сигнал остатка в текущем

интервале обозначим через e (j,i), j=$\overline{0,3}$, $i = \overline{0,39}$, где j=номер группы, i=номер отсчета в группе.

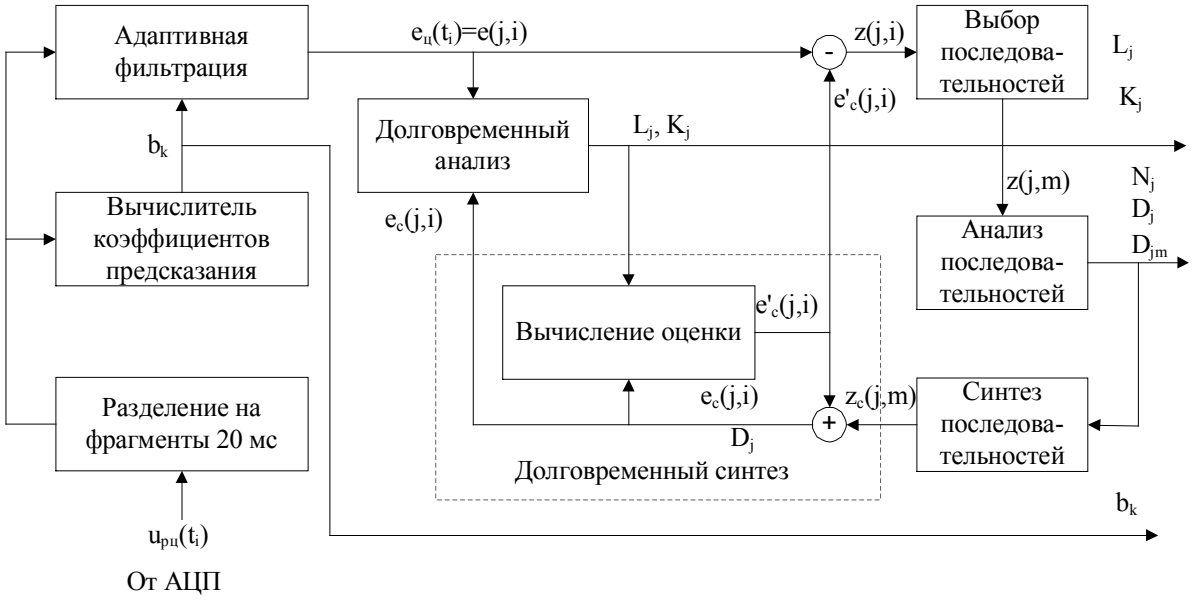

Рис. 5. Речевой кодер

Долговременное предсказание включает анализ и синтез. В результате долговременного синтеза формируется приблизительный остаток предсказания $e_c(j,i) \approx e(j,i)$. Такой же сигнал впоследствии синтезируется в речевом декодере приемника. Рассмотрим выполнение долговременного анализа при условии, что сигнал $e_c(j,i)$ уже сформирован.

В данном сигнале выделяются интервалы, сдвинутые на (-1…-120) отсчетов относительно каждой группы сигнала $e(j,i)$. В этих интервалах проводится поиск 40 отсчетов, наиболее похожих на соответствующую группу. При поиске используется окно шириной 40 отсчетов, которое сдвигается с шагом 1 отсчет. После каждого сдвига вычисляется коэффициент корреляции отсчетов в j группе и в окне r_j. По максимальному коэффициенту $r_{j\,max}$ окно фиксируется и определяется сдвиг между j-й группой и окном, равный L_j. Обозначим отсчеты в таком окне через e_L(j,i). Коэффициенты сглаживания рассчитывается по формуле

$$R_j = \frac{r_{jmax}}{\sum_{i=0}^{39} e_L^2(j,i)} = \sum_{i=0}^{39} e(j,i) \times e_L(j,i) / \sum_{i=0}^{39} e_L^2(j,i) \qquad (6)$$

Значения R_j квантуются четырьмя уравнениями с номерами k_j.

Величины L_j и K_j являются выходными параметрами долговременного предсказания. Сдвиг L_j изменяется в диапазоне (40…120) отсчетов и кодируется 7 битами, номер K_j кодируется 2 битами.

Рассмотрим вычисление оценки при долговременном синтезе. Из сигнала $e_c(j,i)$ с учетом сдвига L_j выбираются те же 40 отсчетов $e_L(j,i)$, которые попали в окно при выполнении долговременного анализа. Эти отсчеты умножаются на квантованный коэффициент $R_{jкв}$, выбираемый по номеру k_j. В результате формируется сигнал

$$e'(j,i) = R_{jкв} \times e_L(j,i) \qquad (7)$$

который является приблизительной копией или оценкой сигнала e(j,i). Погрешность синтеза или остаток долговременной предсказания рассчитывается по формуле

$$z(j,i) = e(j,i) - e'(j,i) = e(j,i - R_{jкв} \times e_L(j,i) \qquad (8)$$

Применение коэффициента сглаживания обеспечивает уменьшение динамического диапазона величин z(j,i) и, соответственно, уменьшение числа битов, требуемых для их кодирования.

Формирование сигнала возбуждения проводится после разделения каждой группы остатка z(j,i), j=$\overline{0,3}$, $i = \overline{0,39}$ на четыре последовательности 13 отсчетов, следующих через z(j,m), где m=$N_j + 3l, l = \overline{0,12}$. Для дальнейшего анализа выбирается одна из четырех последовательностей с максимальной мощностью

$$P_j = \max \sum_m z^2(j,m).$$

Среди 13 отсчетов определяется максимальный отсчет $|z(j,m)_{max}|$, который используется для нормировки

$$B(j,m)=z(j,m)/\ |z(j,m)_{max}| \qquad (9)$$

Далее выполняется квантование. Квантованные отсчеты $|z(j,m)_{max}|_{кв}$ имеют 64 значения с номерами $A_j = 0 \dots 63$, квантованные отсчеты $B(j,m)_{кв}$ имеют 8 значений с номерами $D_{jm} = 0 \dots 7$.

Выходными параметрами являются номера N_j, A_j и D_{jm}, которые кодируются соответственно 2,6 и 3 битами. Эти величины также используются для синтеза последовательностей $z_c(j,m) \approx z(j,m)$. При этом по номерам A_j и D_{jm} выбираются квантованные уровни

$$|z(j,m)_{max}|_{кв} \text{ и } B(j,m)_{кв},$$

После чего вычисляются отсчеты

$$z_c(j,m) = B(j,m)_{кв} \times |z(j,m)_{max}|_{кв} \qquad (10)$$

Временное положение которых задается согласно номером N_j.

Далее выполняется долговременный синтез, при котором отсчеты $z_c(j,m)$ суммируются с оценкой eÈ (j,i):

$$e_c(j,i) = z_c(j,m) + e'(j,i).$$

В результате формируется приблизительный остаток линейного предсказания $e_c(j,i) \approx e(j,i)$. Этот сигнал в следующем интервале 20 мс будет использован при долговременном анализе и при формировании оценки $e'(j,i)$.

Кодирование в первом интервале 20 мс требует ненулевого сигнала $e_c(j,i)$ до начала кодирования. Для этого сразу после включения передатчика на вход кодера подается определенная цифровая последовательность, устанавливающая кодер в требуемое исходное состояние.

В таблице 1 приведен состав выходного слова кодера. Для каждого речевого фрагмента 20 мс слово содержит 260 битов.

Таблица 1

Выходные параметры	Разрядность параметров	Число битов для фрагманта 20 мс
Коэффициенты предсказания b_k	b_1, b_2 – 6 битов, b_3, b_4 – 5 битов, b_5, b_6 – 4 бита, b_7, b_8 – 3 бита	36
Параметры долговременного анализа L_j, K_j(j=0,1,2,3)	L_j – 7 битов, K_j – 2 бита	36
Параметры сигналы возбуждения N_j, A_j, D_{jm}.	N_j – 2 бита, A_j – 6 битов, D_{jm} – 3 бита	188
Длина выходного слова		260

Выходное слово кодера

Рассчитаем характеристики кодера:

- частота формирования слов равна 1 слово/20 мс=50 слов/с
- скорость передачи выходного сигнала 260 бит/20 мс=13 кбит/с
- коэффициент сжатия сигнала АЦП равен 104 кбит/с/13 кбит/с=8

Структурная схема декодера показана на рис. 6.

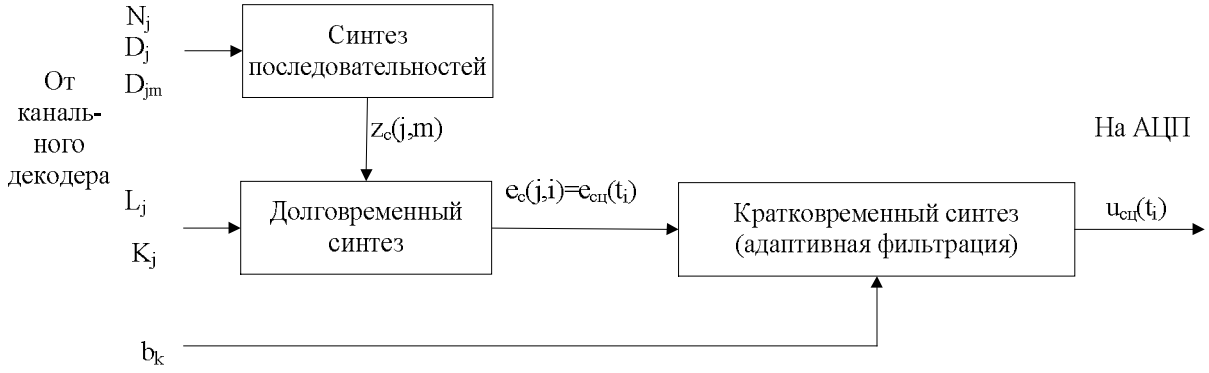

Рис. 6. Речевой кодер

Синтез последовательностей и долговременный анализ выполняются аналогично операциями кодера. Сигнал $e_c(j,i)$ в декодере совпадает с таким же сигналом в кодере (при допущении, что все ошибки исправлены канальным декодером) и служит сигналом возбуждения для синтезирующего фильтра. Фильтр настраивается по коэффициентам b_k. На выходе декодера получают синтезированный цифровой речевой сигнал, подаваемый ЦАП (цифро-аналоговый преобразователь). Погрешность декодирования обусловлена:

1) выбором в кодере 13 отсчетов z(j,m) из 40 отсчетов z(j,i) и потерей остальных 27 отсчетов;

2) квантованием в кодере величин $|z(j,m)_{max}|$ и $B(j,m)$.

Эта погрешность уменьшается за счет операции долговременного предсказания. Вид кодируемого звукового сигнала (речь или шум) не влияет на работу рассмотренного кодека.

Список литературы

1. *В.Г. Потапов, А.Г. Тараненко* Кодирование речи в системах связи с разделением сигналов по форме. Тез. докл. 5 МНТК «Авиа-2003». – Киев, НАУ, 2003.-4с.

2. *М.В. Назаров, Ю.Н. Прохоров* Методы цифровой обработки и передаче речевых сигналов. – М.: Радио и связи, 1985. – 176 с.

Рецензент: Давлет'янц О.І.
Надійшла 21.09.2010