

ТЕОРІЯ ТА МЕТОДИ ОБРОБЛЕННЯ СИГНАЛІВ

УДК 534.782.001:62(045)

О. П. Нечипорук, канд. техн. наук, доц.,
Р. С. Одарченко, канд. техн. наук НАУ, доц.,
В. Г. Потапов, канд. техн. наук, доц.,
А. Г. Тараненко, канд. техн. наук, доц.,
О. П. Ткалич, канд. техн. наук, доц.

ПЕРЕДАЧА РЕЧИ В СИСТЕМАХ ЦИФРОВОЙ СВЯЗИ

Кафедра телекоммуникационных систем Национального авиационного университета
e-mail: tkn@nau.edu.ua

Проанализированы системы передачи речевой информации. Рассмотрены основные модели преобразований речевых сигналов в системах цифровой связи. Приведены основные схемы речевого кодера. Рассчитана погрешность восстановления сигнала для приведенной схемы речевого кодера.

Ключевые слова: цифровая связь, кодирование речи, обработка сигналов, информационные системы, скорость передачи информации, кодер, синтез речевого сигнала.

Введение. Достижения в области цифровой микросхемотехники являются фундаментом практического внедрения методов цифровой передачи речевых сигналов. На основе таких методов созданы системы высококачественной низкоскоростной передачи, которые способны функционировать в реальных условиях и обеспечивают распознавание речи на фоне шумов и помех.

С точки зрения теории информации подобные системы осуществляют кодирование речи. Цель такого преобразования состоит в уменьшении количества битов, отображающих заданное количество информации за определенное время, при этом избыточность цифрового сигнала уменьшается. Одновременное увеличение длительности каждого бита приводит к уменьшению ширины спектра цифрового сигнала, что позволяет использовать канал связи с меньшей полосой пропускания.

Постановка проблемы. Не всегда понятен алгоритм формирования речевых сигналов в системах цифровой связи, который остается недостаточно освященным. В литературе зачастую отсутствуют схемы реальных речевых кодеров, которые широко применяются в системах цифровой связи. Таким образом, необходимо рассмотреть модель цифрового канала, выяснить принципы формирования речевых сигналов в системах цифровой связи, произвести расчет речевого кодера.

Речевое кодирование. В основу теории речевого кодирования положена математическая модель речевого процесса, включающая источник возбуждения, голосовой тракт и акустический излучатель [1]. Широкое применение нашла реализация этой модели с применением линейного предсказания речевого сигнала. По методу линейного предсказания дискретный отсчет речевого сигнала представляется в виде

$$U_p(t_i) = GU_v(t_i) + \sum_{k=1}^m a_k U_p(t_{i-k}), \quad (1)$$

где $U_v(t_i)$ – сигнал возбуждения голосового тракта в текущий момент t_i ; G – коэффициент усиления; a_k – коэффициенты, характеризующие голосовой тракт; m – порядок модели.

Отсчеты $U_p(t_{i-k})$, предшествующие текущему отсчету $U_p(t_i)$, можно измерить и запомнить, а сигнал возбуждения в момент t_i первоначально неизвестен. Поэтому синтезированный речевой отсчет записывают в виде

$$U_c(t_i) = \sum_{k=1}^m b_k U_p(t_{i-k}), \quad (2)$$

где b_k – такие коэффициенты, которые обеспечивают наилучшее качество синтеза речи при неизвестном сигнале возбуждения.

Качество синтеза оценивают мгновенной ошибкой или остатком предсказания

$$e(t_i) = U_p(t_i) - U_c(t_i), \quad (3)$$

при этом нахождение коэффициентов b_k сводится к решению задачи минимизации среднеквадратической ошибки $\sigma = M[e(t_i)]$, где M – это математическое ожидание.

Минимизация ошибки σ достигается при условии

$$\frac{\partial \sigma}{\partial b_k} = 0, \quad k = 1, \dots, m. \quad (4)$$

Выражение (4) представляет собой систему m дифференциальных уравнений с m неизвестными величинами b_k . Коэффициентами при неизвестных величинах в уравнениях (4) служат коэффициенты корреляции речевых отсчетов, а для численного решения системы используются итерационные методы [2].

Принцип анализа и синтеза речевого сигнала поясняет рис. 1.

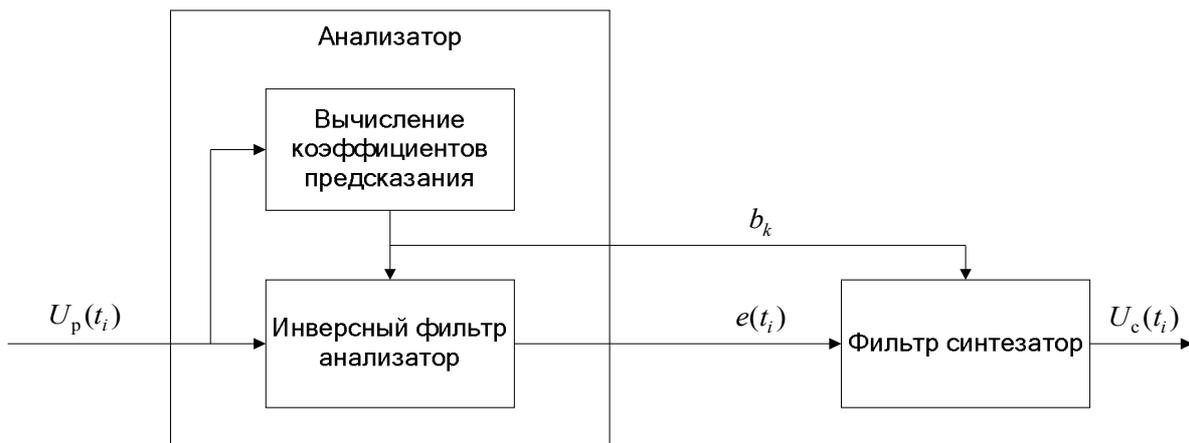


Рис. 1. Анализ и синтез речевого сигнала

По коэффициентам b_k , предварительно вычисленным в анализаторе, настраивается передаточная характеристика фильтра-синтезатора. Этот фильтр моделирует голосовой тракт и синтезирует речевые отсчеты из сигнала возбуждения. Покажем, что в качестве такого сигнала можно использовать остаток предсказания.

Коэффициенты b_k являются оптимальными оценками коэффициентов a_k в формуле (1), т. е. $b_k = a_{k \text{ opt}}$ при неизвестном сигнале возбуждения. После подстановки в формулу (1) оптимальных значений $a_k = a_{k \text{ opt}} = b_k$ по формуле (3) получаем

$$e(t_i) = U_p(t_i) - U_c(t_i) = GU_v(t_i),$$

где $U_c(t_i)$ определяется формулой (2).

Следовательно, остаток предсказания является возбуждающим сигналом, т. е. входным сигналом синтезатора. Для получения этого остатка речевой сигнал в анализаторе необходимо пропустить через инверсный фильтр, который имеет передаточную характеристику, обратную характеристике синтезирующего фильтра. Инверсный фильтр

также настраивается по коэффициентам b_k . При выводе формул для передаточных характеристик фильтров используется математический аппарат Z-преобразований [1].

Рассмотренный метод анализа и синтеза применяется для обработки случайных стационарных сигналов. Реальный речевой сигнал является случайным нестационарным, поскольку его вероятностные характеристики изменяются во времени. В этом случае линейное предсказание выполняется на интервалах квазистационарности Δt , в которых вероятностные характеристики сигнала приблизительно постоянны. На практике речевой сигнал в анализаторе разделяется на фрагменты длительностью. В каждом интервале вычисляются новые коэффициенты b_k , по которым перестраиваются адаптивные фильтры анализатора и синтезатора.

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

В кодеке реализован метод RPE-LPT (Regular Pulse Excitation Long Term Prediction – долговременное предсказание с регулярным импульсным возбуждением). Структурная схема кодера показана на рис. 2.

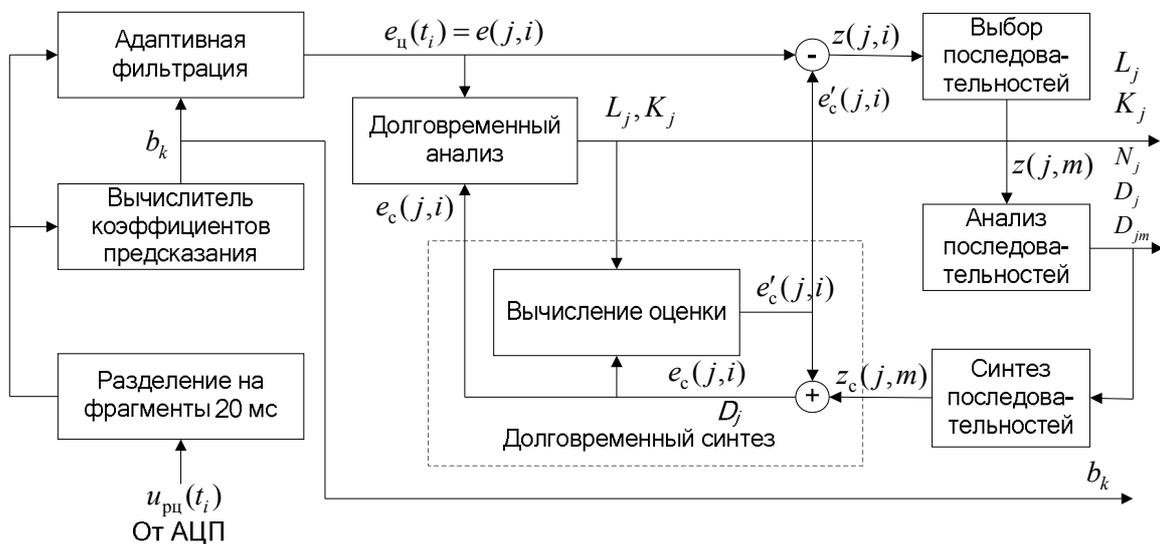


Рис. 2. Речевой кодер

На вход кодера от аналого-цифрового преобразователя (АЦП) поступает цифровой сигнал в виде двоичных 13-разрядных отсчетов со скоростью передачи 104 кбит/с. Кодирование включает три 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.

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 $(-1, \dots, -120)$ 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 в 1 отсчет. После каждого сдвига вычисляется коэффициент корреляции отсчетов в j -й 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 отсчетов. Обозначим отсчеты в таком окне через $e_L(j,i)$. Коэффициент сглаживания рассчитывается по формуле

$$R_j = \frac{r_{j\max}}{\sum_{i=0}^{39} e_L^2(j,i)} = \frac{\sum_{i=0}^{39} e(j,i)e_L(j,i)}{\sum_{i=0}^{39} e_L^2(j,i)}.$$

Значения R_j квантуются четырьмя уровнями 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 $R_{j\text{кв}}$, выбираемый по номеру K_j . В результате формируется сигнал

$$e'(j,i) = R_{j\text{кв}} e_L(j,i),$$

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

$$Z(j,i) = e(j,i) - e'(j,i) = e(j,i) - R_{j\text{кв}} e_L(j,i).$$

Коэффициент сглаживания обеспечивает уменьшение динамического диапазона величины $Z(j,i)$ и соответственно уменьшение количества битов, требуемых для их кодирования.

Сигнал возбуждения формируется после деления каждой группы остатка $Z(j,i)$, $j = \overline{0,3}$, $i = \overline{0,39}$ на четыре последовательности по 13 отсчетов, следующих через равные промежутки времени (регулярно).

Последовательность с номером N_j обозначим через $Z(j,m)$, где $m = N_j + 3l$, $l = \overline{0,21}$. Для дальнейшего анализа выбирается одна из четырех последовательностей с максимальной мощностью $P_j = \max_m \sum Z^2(j,m)$. Среди 13 отсчетов определяется максимальный отсчет $|Z(j,m)_{\max}|$, который используется для нормирования

$$B(j,m) = \frac{Z(j,m)}{|Z(j,m)_{\max}|}.$$

Далее выполняется квантование. Квантованные отсчеты $\left| Z(j, m) \right|_{\max_{\text{КВ}}}$ имеют 64 значения с номерами $A_j = 0 \dots 63$, квантованные отсчеты $B(j, m)_{\text{КВ}}$ 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 $\left| Z(j, m) \right|_{\max_{\text{КВ}}}$ и $B(j, m)_{\text{КВ}}$, после чего вычисляются отсчеты

$$Z_c(j, m) = B(j, m)_{\text{КВ}} \cdot \left| Z(j, m) \right|_{\max_{\text{КВ}}},$$

временное положение которых задается согласно номеру N_j .

Далее выполняется долговременный синтез, при котором отсчеты $Z_c(j, m)$ суммируются с оценкой $e'(j, i)$:

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

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.

Состав выходного слова кодера приведен в таблице. В каждом речевом фрагменте 20 мс слово содержит 260 битов.

Выходное слово кодера

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

Рассчитаем характеристики кодера:

– частота формирования слов равна $\frac{1 \text{ слово}}{20 \text{ мс}} = 50 \frac{\text{слов}}{\text{с}}$;

– скорость передачи выходного сигнала равна $\frac{260 \text{ бит}}{20 \text{ мс}} = 13 \frac{\text{кбит}}{\text{с}}$;

– коэффициент сжатия сигнала АЦП равен $\frac{104 \frac{\text{кбит}}{\text{с}}}{13 \frac{\text{кбит}}{\text{с}}} = 8$.

Структурная схема декодера показана на рис. 3.

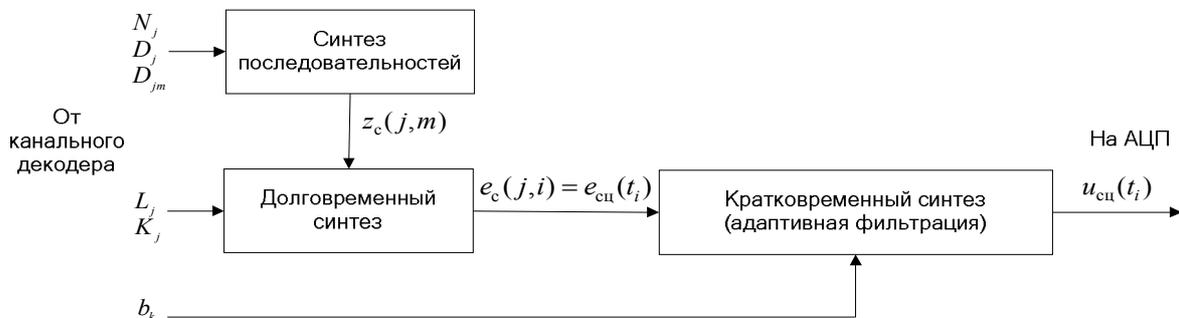


Рис. 3. Речевой кодер

Синтез последовательностей и долговременный синтез выполняются аналогично операциям кодера. Сигнал $e_c(j, i)$ в декодере совпадает с таким же сигналом в кодере (при допущении, что все ошибки исправлены каналным декодером) и служит сигналом возбуждения для синтезирующего фильтра. Фильтр настраивается по коэффициентам b_k . На выходе декодера получают синтезированный речевой сигнал, подаваемый на цифро-аналоговый преобразователь. Вид кодируемого звукового сигнала (речи или шума) не влияет на работу рассмотренного кодера.

Выводы. Анализ рассмотренного кодера показывает, что погрешность декодирования обусловлена двумя причинами: выбором в кодере 13 отсчетов $Z(j, m)$ из 40 отсчетов $Z(j, i)$ и потерей остальных 27 отсчетов, а также квантованием в кодере величин $|Z(j, m)_{\max}|$ и $B(j, m)$.

Уменьшение этой погрешности достигается применением операции долговременного предсказания, что, в свою очередь, способствует повышению синтезированной речи.

Список литературы

1. Рабинер Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер – М.: Радио и связь, 1981. – 496 с.
2. Назаров М. В. Методы цифровой обработки и передачи речевых сигналов / М. В. Назаров, Ю. М. Прохоров – М.: Радио и связь, 1985. – 176 с.

Р. С. Одарченко, В. Г. Потапов, А. Г. Тараненко, О. П. Ткаліч

Передавання мови в системах цифрового зв'язку

Проаналізовано системи передавання мовної інформації. Розглянуто основні моделі перетворень мовних сигналів у системах цифрового зв'язку. Наведено основні схеми мовного кодера. Розраховано похибку відновлення сигналу для наведеної схеми мовного кодера.

R. S. Odarchenko, V. G. Potapov, A. G. Taranenko, O. P. Tkalich

Speech transfer in digital communication systems

Paper is devoted to analysis of the existing speech information transmission systems. The basic model transforms voice signals in digital communication systems. The basic scheme of speech coder. Signal error for used speech coder recovery scheme is calculated.