Лекция 7
СТАНДАРТ CELP

Рассмотрим стандарт CELP, который является основой всех стандартных речевых кодеров типа анализ-через-синтез, использующих принцип линейного предсказания (G.729, G.723.1, G.728, IS-54, IS-96, RPE-LTP (GSM), FS-1016 (CELP) и т. д.) CELP – это аббревиатура от *C*ode *E*xcited *L*inear *P*redictive coder (кодер на основе линейного предсказания, возбуждаемый кодом). Этот кодер обрабатывает речевой сигнал кадрами. В кодере используется линейный предсказывающий фильтр 10-го порядка, выполняющий кратковременное (short-term) предсказание формант речевого сигнала на каждом кадре. Так называемое долговременное (long-term) предсказание периода основного тона (pitch) выполняется с помощью адаптивной кодовой книги. Разность между входным речевым сигналом и фильтрованным кодовым словом из адаптивной книги векторно квантуется с помощью стохастической кодовой книги. Оптимальные в смысле минимума перцептуально взвешенной среднеквадратической ошибки взвешенные кодовые слова из адаптивной и стохастической книги задаются индексом (номером в кодовой книге) и коэффициентом усиления (взвешивающим коэффициентом). Перцептуальное взвешивание среднеквадратической ошибки используется для улучшения субъективного восприятия синтезированного речевого сигнала.

Вычислительная сложность стандарта CELP в основном определяется сложностью поиска в адаптивной и стохастической кодовых книгах. Поиск в этих книгах выполняется полным перебором. Кроме того, поскольку нахождение входного сигнала фильтра (ошибки предсказания) соответствует неортогональному преобразованию (матрица преобразования содержит сдвиги импульсной характеристики фильтра), вычисление ошибки производится не в области сигналов возбуждения, а в области синтезированных речевых сигналов. Отсюда и название этого класса кодеров – кодеры с анализом-через-синтез. Другими словами, для того, чтобы найти оптимальное кодовое слово из адаптивной или стохастической книги необходимо профильтровать каждое слово из этой книги линейным предсказывающим фильтром, вычислить взвешенную среднеквадратическую ошибку между целевым сигналом и взвешенным профильтрованным словом и найти слово, соответствующее минимуму ошибки. Таким образом, вычислительная сложность и качество синтезированной речи существенным образом зависят от размеров кодовых книг.

Стандарт CELP использует входные речевые сигналы дискретизованные с частотой 8 кГц, каждый отсчет представляет собой 16-битовое целое число. Речевой сигнал обрабатывается кадрами по 240 отсчетов, т.е. 30 мс, каждый. Кадр разбивается на четыре подкадра по 60 отсчетов, т.е. длительностью 7.5 мс. Анализ речевого сигнала, выполняемый в кодере, включает три основных операции: 1) кратковременное предсказание или построение предсказывающего фильтра, 2) долговременное предсказание или поиск в адаптивной кодовой книге, и 3) обновляющий поиск в стохастической кодовой книге.

Сжатый файл, получаемый на выходе кодера содержит номер слова из стохастической книги и его проквантованный взвешивающий коэффициент, номер слова из адаптивной книги и его проквантованный взвешивающий коэффициент и 10 проквантованных ЛСП.

Кодер стандарта CELP показан на Рис. 7.1. Аппроксимирующий вектор $\hat{\mathbf{s}}$ вычитается из входного речевого сигнала, и квадрат разности затем перцептуально взвешивается. Эта перцептуально взвешенная ошибка управляет поиском в адаптивной

кодовой книге, выполняемом по принципу “анализ-через-синтез” или “с замкнутой петлей” (closed-loop analysis). Процедура поиска находит индексы и коэффициенты усиления i_a (g_a) и i_s(g_s), минимизирующие перцептуально взвешенную ошибку, в адаптивной и стохастической книгах, соответственно. Линейный предсказывающий фильтр вычисляется по методу Левинсона-Дарбина или так называемым методом “анализа с открытой петлей” (open-loop analysis).

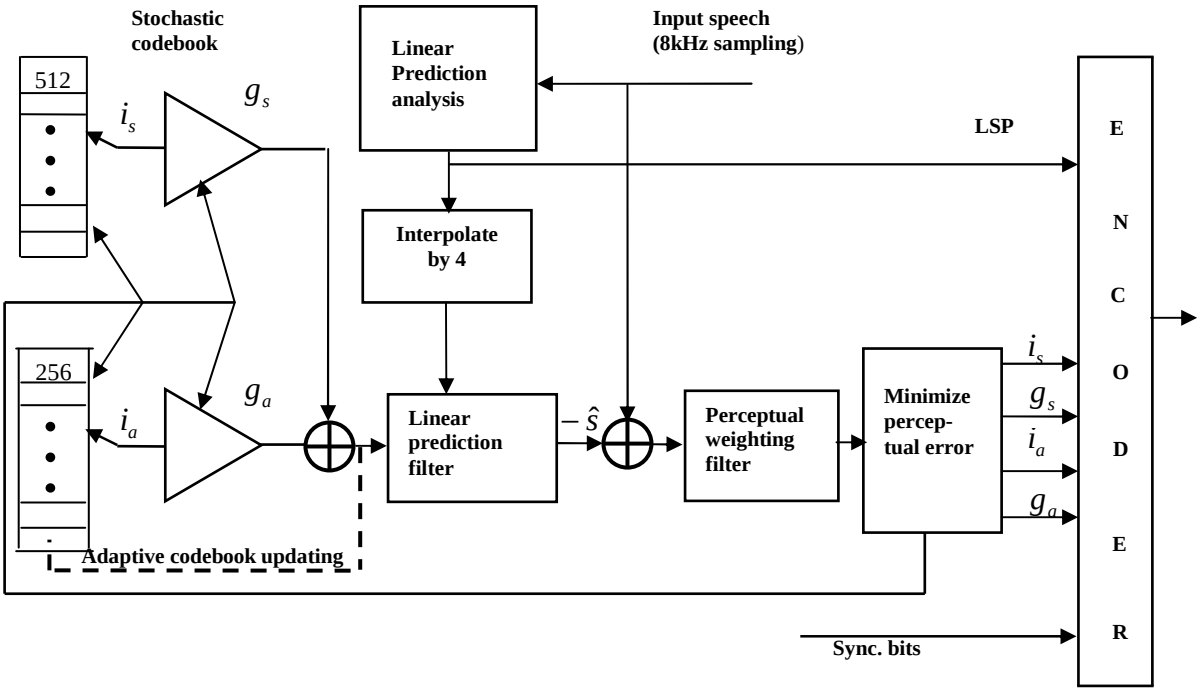

Рис. 7.1 Кодер стандарта CELP

Коэффициенты предсказывающего фильтра вычисляются один раз на кадр (30мс) с использованием окна Хэмминга. Фильтр линейного предсказания имеет вид $1/A(z)$. Перцептуальный взвешивающий фильтр с передаточной функцией $A(z)/A(z/\gamma)$, $\gamma = 0.8$ расширяет полосу частот исходного фильтра делением z на коэффициент γ, результирующая передаточная функция перцептуально взвешенного фильтра имеет вид $\frac{1}{A(z/\gamma)}$. После того как коэффициенты фильтра вычислены по методу Левинсона-Дарбина, вычисляются линейные спектральные параметры. Таким образом, предсказывающий фильтр передается десятью ЛСП, которые неравномерно скалярно квантуются в соответствии с Табл. 7.1. Общее число битов, затрачиваемых на передачу ЛСП равно 34. Так как ЛСП передаются только один раз на кадр, а используются на каждом подкадре, то они линейно интерполируются между текущим и будущим кадрами. Вычисление фильтра на будущем кадре предполагает так называемое “заглядывание вперед”.

Как уже было сказано выше, поиск в адаптивной книге выполняется по методу “анализа с замкнутой петлей” с использованием в качестве критерия модифицированной среднеквадратической ошибки. Номер одного из 256 кодовых слов в книге кодируется восемью битами. Для уменьшения вычислительной сложности кодер может использовать для поиска любое подмножество кодовых слов из адаптивной кодовой книги. Поиск в адаптивной кодовой книге служит для предсказания периодической части возбуждения фильтра (ошибки предсказания). В

книге хранится предыдущий сигнал возбуждения с задержкой M в диапазоне от 20 до 147 отсчетов. Для кадра речевого сигнала Рис.7.3, соответствующий сигнал возбуждения показан на Рис. 7.2. Каждое кодовое слово в книге представляет собой 60 отсчетов предыдущего сигнала возбуждения, задержанного на M отсчетов. На каждом подкадре производится поиск среди 256 кодовых слов, соответствующих 128 целым и 128 нецелым задержкам $20 \le M \le 147$. Для задержек меньших по величине, чем длина подкадра ($M < 60$) кодовые слова содержат M начальных отсчетов предыдущего сигнала возбуждения. Для того, чтобы сформировать кодовое слово длины 60 отсчетов, начальные отсчеты некоторым образом повторяются.

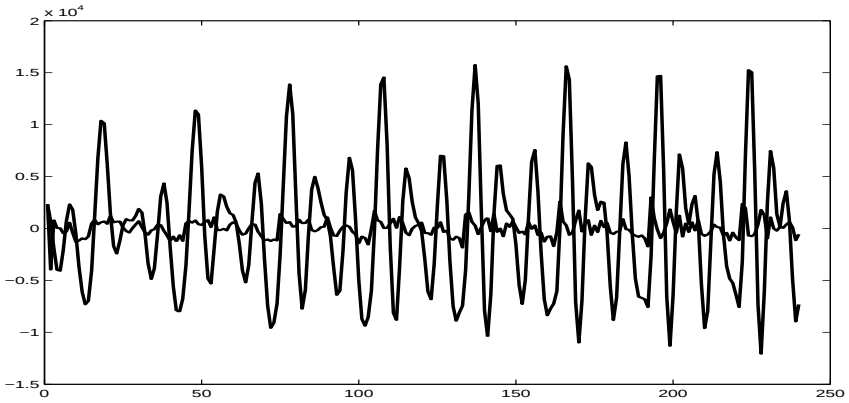

Рис.7.2 Сигнал возбуждения фильтра линейного предсказания

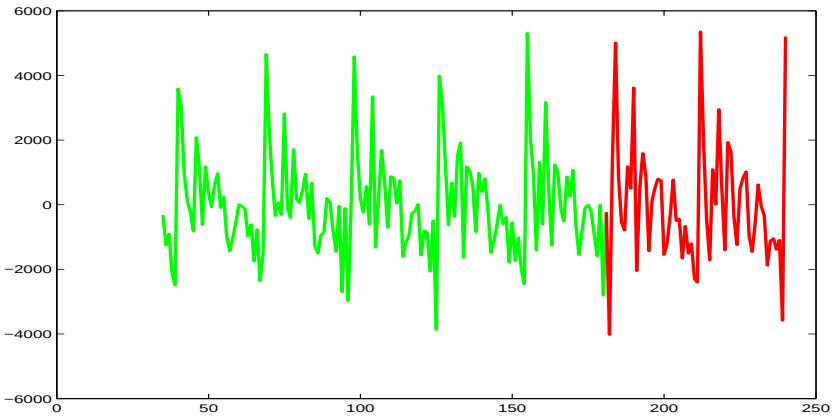

Рис. 7.3 Адаптивная книга и сигнал возбуждения текущего подкадра

Кодовые слова адаптивной кодовой книги выглядят следующим образом:

Таблица 7.1

Номер	Задержка	Номера отсчетов в кодовом слове
255	147	-147, -146, -145,…, -89,-88
254	146	-146, -145, -144,…, -88,-87
253	145	-145, -144, -143,…, -87,-86
…	…	…
131	61	-61, -60, -59, …, -2, -1
…	…	…
1	21	-21, …,-2,-21,…,-2,-21,…, -2
0	20	-20,…, -1, -20,…-1, -20,…, -1

Для нахождения оптимального слова в адаптивной кодовой книге используется линейное предсказание 1-го порядка. Пусть $\mathbf{s}$ представляет собой входной речевой вектор и $\mathbf{a}_i$ - это фильтрованное кодовое слово $\mathbf{c}_i$ из адаптивной кодовой книги, тогда мы ищем

$$\min_i \|\mathbf{s} - g_a \mathbf{a}_i\| = \min_i \left\{ \|\mathbf{s}\|^2 - 2g_a(\mathbf{s}, \mathbf{a}_i) + g_a^2 \|\mathbf{a}_i\|^2 \right\}, \qquad (10.1)$$

где g_a - это оптимальный коэффициент предсказания, называемый коэффициентом усиления (adaptive codebook gain).

Беря производную от правой части в (10.1) по g_a и приравнивая ее нулю, находим, что оптимальный коэффициент усиления равен

$$g_a = \frac{(\mathbf{s}, \mathbf{a}_i)}{\|\mathbf{a}_i\|^2}.$$

Другими словами оптимальный взвешивающий коэффициент представляет собой отношение взаимной корреляционной функции целевого сигнала и фильтрованного кодового слова к энергии фильтрованного кодового слова. Подставляя оптимальное значение взвешивающего коэффициента в (10.1), получаем

$$\min_i \|\mathbf{s} - g_a \mathbf{a}_i\| = \min_i \left\{ \|\mathbf{s}\|^2 - \frac{2(\mathbf{s}, \mathbf{a}_i)^2}{\|\mathbf{a}_i\|^2} + \frac{(\mathbf{s}, \mathbf{a}_i)^2}{\|\mathbf{a}_i\|^2} \right\} = \min_i \left\{ \|\mathbf{s}\|^2 - \frac{(\mathbf{s}, \mathbf{a}_i)^2}{\|\mathbf{a}_i\|^2} \right\}. \qquad (10.2)$$

Минимизация (10.2) по i эквивалентна максимизации последнего члена в (10.2) так как первый член не зависит от кодового слова $\mathbf{a}_i$. Таким образом, процедура поиска в адаптивной кодовой книге находит кодовое слово $\mathbf{c}_i$, которое максимизирует целевую функцию (match function) m_i:

$$m_i = \frac{(\mathbf{s}, \mathbf{a}_i)^2}{\|\mathbf{a}_i\|}.$$

Номер слова из адаптивной книги i_a и коэффициент g_a передаются четыре раза на кадр (каждые 7.5 мс). Коэффициент кодируется в диапазоне от –1 до +2 с использованием неравномерного скалярного квантования со скоростью 5 бит/отсчет.

Процедуры поиска в стохастической и адаптивной книгах внешне идентичны. Они различаются только кодовыми книгами и целевыми векторами. Для уменьшения вычислительной сложности поиск в адаптивной и стохастической книгах выполняется в две ступени. После того, как найдено оптимальное кодовое слово в адаптивной книге, вычисляется разность между входным вектором $\mathbf{s}$ и взвешенным оптимальным фильтрованным кодовым словом, т. е.,

$$\mathbf{u} = \mathbf{s} - g_a \mathbf{a}_{opt}.$$

Поиск в стохастической кодовой книге также выполняется по методу “анализа с замкнутой петлей” с использованием критерия перцептуально взвешенной среднеквадратической ошибки. Мы находим кодовое слово $\mathbf{x}_i$, которое максимизирует следующую целевую функцию

$$\frac{(\mathbf{u},\mathbf{y}_i)^2}{\|\mathbf{y}_i\|^2},$$

где $\mathbf{y}_i$ - это фильтрованное кодовое слово $\mathbf{x}_i$.

Девять битов отводится на кодирование номера слова в стохастической кодовой книге, содержащей 512 кодовых слов. Для уменьшения вычислительной сложности возможен поиск в любом подмножестве этой книги. Номер слова и взвешивающий коэффициент передаются четыре раза на кадр. Коэффициент (положительный и отрицательный) квантуется скалярным неравномерным квантователем со скоростью 5 бит/отсчет. Стохастическая книга содержит псевдослучайные последовательности длиной по 60 отсчетов квантованные на три уровня (–1,0,+1) .

Общее число битов на кадр можно вычислить как

$$4(b_{g_s} + b_{i_s} + b_{g_a} + b_{i_a}) + b_{LSP} = 4(5+8+5+9)+34 = 142,$$

где b_{g_s}, b_{i_s} и b_{g_a}, b_{i_a} - это число битов на номер и коэффициент в стохастической и адаптивной книгах, соответственно, b_{LSP} - число битов на ЛСП. Принимая во внимание, что длительность кадра равна 30 мс, мы получаем, что битовая скорость в стандарте CELP примерно равна

$$R = \frac{142}{30 \cdot 10^{-3}} \approx 4733 \text{ бит/с.}$$

Добавляя биты на синхронизацию и исправление ошибок, получаем, что битовая скорость равна 4800 бит/с.