

РАЗРАБОТКА ШИРОКОПОЛОСНОГО КОДЕРА РЕЧИ

Описывается усовершенствованный подход для повышения качества в методе субполосного CELP-кодирования. Предлагается модель широкополосного CELP-кодера с реконфигурацией структуры кодовой книги на основании оценок порога маскирования, полученного из психоакустической модели, что позволяет снизить объем передаваемой информации, сохраняя высокое качество.

Линейное предсказание с мультикодовым управлением (англ. Code Excited Linear Prediction, CELP) представляет собой алгоритм кодирования речи, обеспечивающий лучшее качество, чем алгоритмы с низкой скоростью прохождения битов информации, такие, как аудиокодеки RELP и LPC. Наряду с вариантами, как, например, ACELP, RCELP, LD-CELP и VSELP, на сегодня это наиболее широко используемый алгоритм кодирования речи [1].

Кодеры речевых сигналов, использующие алгоритм CELP, относятся к классу гибридных и занимают промежуточное положение между кодерами формы, в которых сохраняется форма колебания речевого сигнала в процессе его дискретизации и квантования, и параметрическими вокодерами, основанными на процедурах оценки и кодирования небольшого числа параметров речи. При этом в кодерах CELP сохраняются преимущества обоих типов кодеров.

Метод кодирования CELP основан на линейной авторегрессионной модели процесса формирования и восприятия речи. Он обеспечивает более высокое качество декодированной речи, устойчивость к входному акустическому шуму и ошибкам в канале связи по сравнению с иными принципами кодирования. Наиболее эффективно применение этого метода при передаче речевого сигнала в диапазоне скоростей от 14 до 16 Кбит/с [2].

Предлагается следующая структурная схема широкополосного CELP-кодера (рис.1), ядром которой является мультиполосная кодовая книга, состоящая из субполосных кодовых книг форм и коэффициентов усиления. Входной сигнал квантуется при помощи 16 бит/отсчет с частотой

дискретизации $f_s=16\text{кГц}$, после чего кодируется широкополосным кодером, структура которого представлена на рис.1.

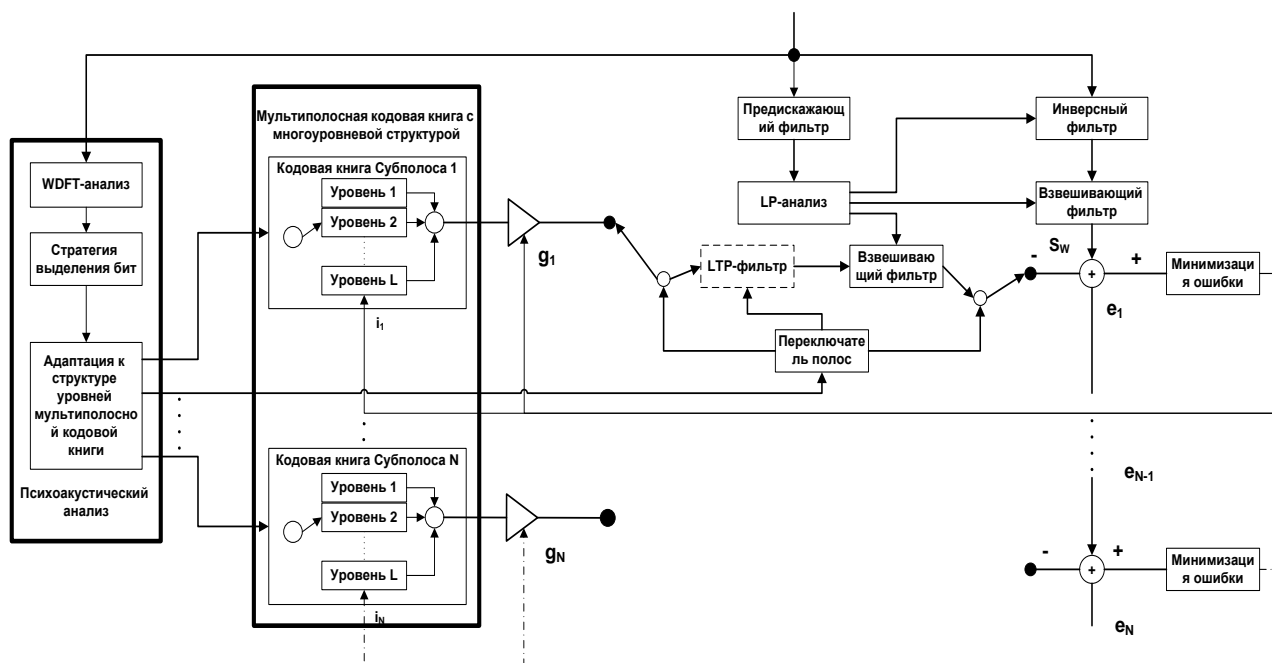


Рисунок 1. Структура широкополосного кодера

Кодер использует блочный LP-анализ, без интерполяции коэффициентов линейного предсказания. В качестве наиболее оптимального для полосы частот 50 – 8000 Гц, был определен $p=16$ порядок фильтра-предсказателя (Short-Term Predictor - STP). Для улучшения работоспособности взвешивающего фильтра используется предсказывающий фильтр.

Оценка порога маскирования основана на общей психоакустической модели [3]. Данная процедура была подстроена под широкополосную речь с полосой 8 кГц и модифицирована для использования свойств неравномерного частотного разрешения (WDFT). Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости [4].

По вычисленной энергии порога маскирования осуществляется определение важности кодируемых субполос:

$$SPE_b = \log_2 \left(\frac{E_b}{\sqrt{\frac{6 \cdot T_b}{\text{bandwidth}_b}}} + 1 \right), \text{ [бит/полосу]} \quad (1)$$

где E_b - энергия субполосы, T_b - субполосный порог маскирования, bandwidth_b - ширина субполосы b в спектральных отсчетах.

Принцип изменения структуры кодовой книги заключается в следующем.

Оригинальный речевой сигнал (очередной кодируемый фрейм) поступает на вход блока WDFT-анализа, в котором осуществляется оценка субполосных порогов маскирования T_b , затем производится оценка перцептуальной энтропии каждой из субполос кодера SPE_b . Таким образом, на выходе мы имеем своего рода стратегию выделения бит или вектор SPE , содержащий значения важности кодируемых субполос для текущего фрейма сигнала.

Таким образом, можно сделать вывод о том, что предлагаемая схема структуры мультиполосной многоуровневой кодовой книги кодера позволяет существенно снизить вклад данных об индексах векторов возбуждения и коэффициентах их усиления в результирующий информационный поток. При этом увеличение количества бит, необходимых для кодирования информации, наблюдается лишь в значимых областях (фрагментах) речевого сигнала.

Перечень литературы:

1. <http://ru.wikipedia.org/wiki/CELP>
2. http://www.sagatelecom.ru/encyclopedia/protocol/detail.php?SECTION_ID=28&ID=99
3. <http://www.compression-pointers.ru>
4. Цифровые банки фильтров: анализ, синтез и применение в мультимедиа системах: / Петровский А.А., Парфенюк М., Лившиц М.З. Изд-во БГУИР 2006. - 67 с.