УДК 621.391

ЦИФРОВОЕ КОДИРОВАНИЕ ШИРОКОПОЛОСНОГО РЕЧЕВОГО СИГНАЛА В ЗАДАЧЕ ТЕЛЕФОНИИ

Рыболовлев А.А., *к.т.н., Академия Федеральной службы охраны Российской Федерации г. Орёл, e-mail: rybolovlev@rambler.ru*

DIGITAL CODING OF WIDEBAND SPEECH SIGNAL ON TELEPHONY TASK

Rybolovlev A.A.

The article describes the adaptive multirate wideband (AMR-WB) speech codec, which was used by the International Telecommunication Union – Telecommunication Sector (ITU-T) for wideband speech coding around 16 kbit/s (Recommendation G.722.2). AMR-WB uses an extended speech bandwidth from 50 Hz to 7 kHz and gives high speech quality and voice naturalness. Codec operates at a multitude of bit rates ranging from 6,6 kbit/s to 23,85 kbit/s. The bit rate may be changed at any 20-ms frame boundary. The paper details AMR-WB algorithmic description.

Key words: speech signal, speech coding, codebook, linear prediction coefficients.

Ключевые слова: речевой сигнал, кодирование речи, кодовая книга, коэффициенты линейного предсказания.

Представлен адаптивный многоскоростной кодек широкополосного речевого сигнала (AMR-WB), использованный Сектором стандартизации Международного Союза электросвязи для кодирования широкополосного речевого сигнала со средней скоростью около 16 кбит/с (Рекомендация G.722.2). AMR-WB кодирует речь в диапазоне частот от 50 Гц до 7 кГц и обеспечивает высокое качество и натуральность речевого сигнала. Кодек функционирует на нескольких скоростях кодирования от 6,6 кбит/с до 23,85 кбит/с. Скорость кодирования может быть изменена на любом кадре длительностью 20 мс. Статья детализирует алгоритмическое описание AMR-WB.

Введение

Частотное ограничение речевого сигнала (РС) в задаче телефонии диапазоном 300 – 3400 Гц исторически было обусловлено необходимостью экономии частотно-временного ресурса, предназначенного для передачи РС на расстояние в режиме реального времени при сохранении требуемого уровня разборчивости речи. Используемый частотный диапазон в этом случае значительно ограничивает энергетику отдельных групп звуков, например – фрикативных согласных, точность передачи параметров основного тона и формантных областей РС [1, 2], что объективно ухудшает условия восприятия телефонной речи.

Современный этап развития цифровых инфокоммуникационных технологий и достигнутый уровень производительности микропроцессоров делают возможным передачу по телефонным сетям речевого сигнала с более широким частотным диапазоном. Целью такого подхода является повышение субъективного качества телефонной (как правило – синтетической) речи без критического роста скорости кодирования и временно́й задержки на обработку сигнала. При этом под повышением качества речи подразумевается не только выполнение требований по разборчивости, но и улучшение показателей её натуральности (естественности) и узнаваемости говорящего при согласовании со спектральными характеристиками слухового аппарата [3].

Исследования в направлении расширения частотного диапазона телефонной речи привели к появлению ряда алгоритмов кодирования широкополосного речевого сигнала (ШРС), рассчитанных на использование широкого (WB – Wideband, до 8 кГц), сверхширокого (SWB – Super wideband, до 16 кГц) и полного (FB – Fullband, до 20 кГц) частотных диапазонов [4, 5]. В статье представлен адаптивный многоскоростной кодек широкополосного речевого сигнала (AMR-WB – Adaptive Multi-Rate Wideband) в варианте, изложенном в Рекомендации МСЭ-Т G.722.2 «Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband».

Общая характеристика кодека

Кодек реализует технологию гибридного кодирования речи на основе метода линейного предсказания с возбуждением от алгебраического кода (ACELP) [6]. Структурные схемы кодера и декодера со структурой кодового слова представлены на рис. 1 и рис. 2 соответственно. Особенности процедур обработки речевого сигнала заключаются в следующем:

– обработка и передача ШРС с диапазоном частот от 50 Гц до 7 кГц;

– предварительная дискретизация ШРС с частотой 16 кГц и представление в формате 14-битной ИКМ, что формирует входной цифровой поток со скоростью передачи 224 кбит/с;

– разделение ШРС на нижнюю (50 Гц – 6400 Гц) и верхнюю (6400 Гц – 7000 Гц) полосы частот с последующим применением раздельных процедур (трактов) их обработки;

– использование процедуры предыскажения с целью относительного выравнивания по мощности спектральных составляющих обрабатываемого ШРС в диапазоне частот 50 – 6400 Гц (обеливание речевого сигнала);

– использование традиционной локально-стационарной модели речевого сигнала с длительностью кадра в

20 миллисекунд, при этом кадр входного ИКМ-потока содержит 320 отсчетов дискретного ШРС, представленных 4480 битами;

– применение режима прерывистой передачи (DTX – Discontinuous Transmission) на основе использования детектора активности речи (VAD – Voice Activity Detector); выходной цифровой поток кодера в режиме паузы содержит информацию о фоновом шуме, необходимую для функционирования генератора комфортного шума декодера, и имеет скорость 1,75 кбит/с;

– моделирование сигнала ошибки предсказания (сигнала возбуждения) суммой масштабированных сигналов (векторов) возбуждения адаптивной и фиксированной кодовых книг;

– разделение кадра ШРС на четыре подкадра длительностью 5 мс с целью определения параметров сигнала возбуждения синтезирующего фильтра линейного предсказания;

– реализация процедуры кратковременного линейного предсказания (STP – Short Term Prediction) с расчетом линейных спектральных пар (ЛСП) на длительности кадра;

– реализация процедуры долговременного линейного предсказания (LTP – Long Term Prediction) на основе определения периода основного тона (ОТ) и его использования в адаптивной кодовой книге;

– структурно-параметрическая степень адаптации;

– использование девяти скоростных режимов кодирования активной речи: 23,85, 23,05, 19,85, 18,25, 15,85, 14,25, 12,65, 8,85 и 6,6 кбит/с; кодовое слово, отображающее кадр речевого сигнала, составляет при этом 477, 461, 397, 365, 317, 285, 253, 177 и 132 бита соответственно; детализация распределения бит кодового слова по кодируемым параметрам текущего кадра ШРС для различных скоростей кодирования представлена в табл. 1; адаптивный выбор режимов кодирования используется для перераспределения информационных ресурсов между процедурами кодирования источника и канального кодирования в зависимости от состояния канала связи и в данной статье не рассматривается;

– активное использование при обработке ШРС стандартных процедур цифровой обработки сигналов (децимации, интерполяции, оконного взвешивания, фильтрации и др.).

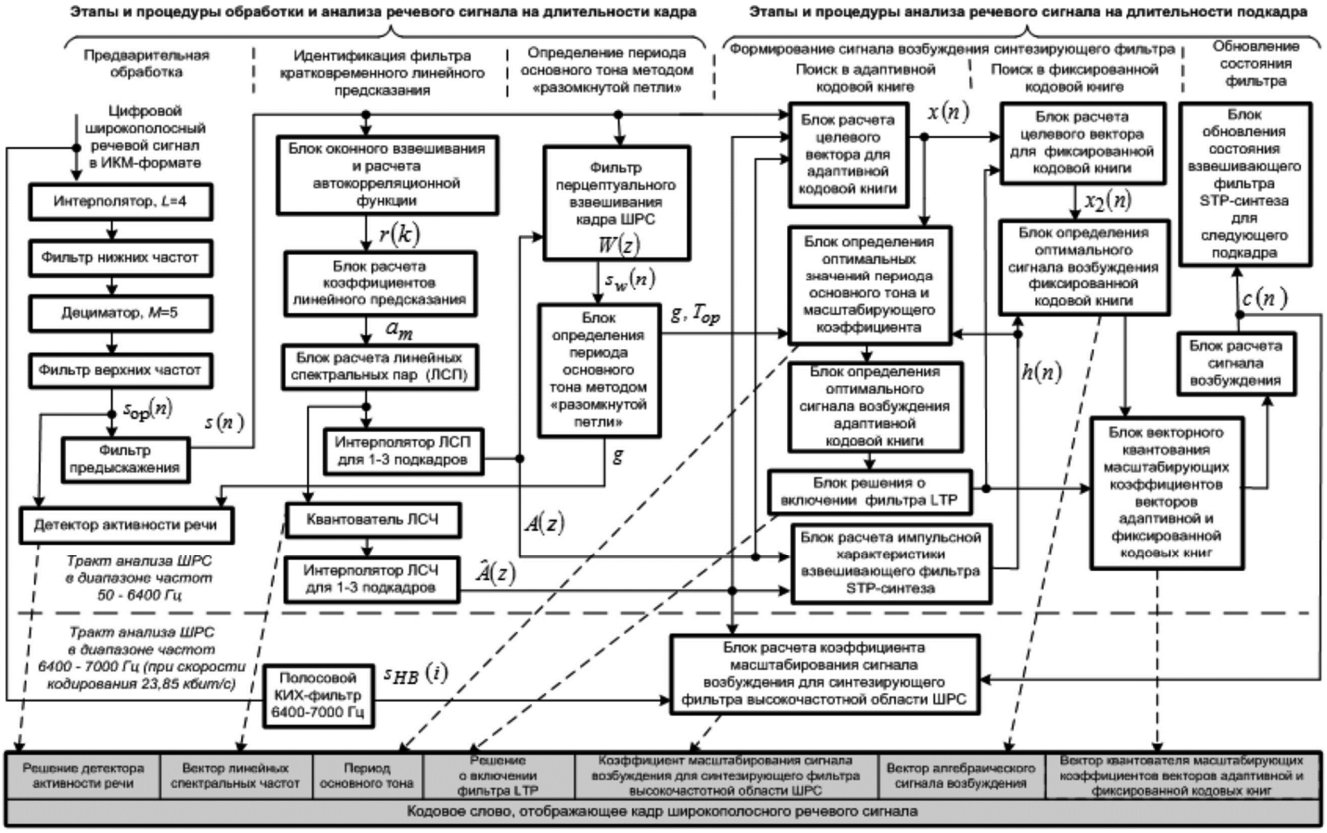

Рис. 1. Структурная схема кодера AMR-WB и структура кодового слова, отображающего кадр широкополосного речевого сигнала

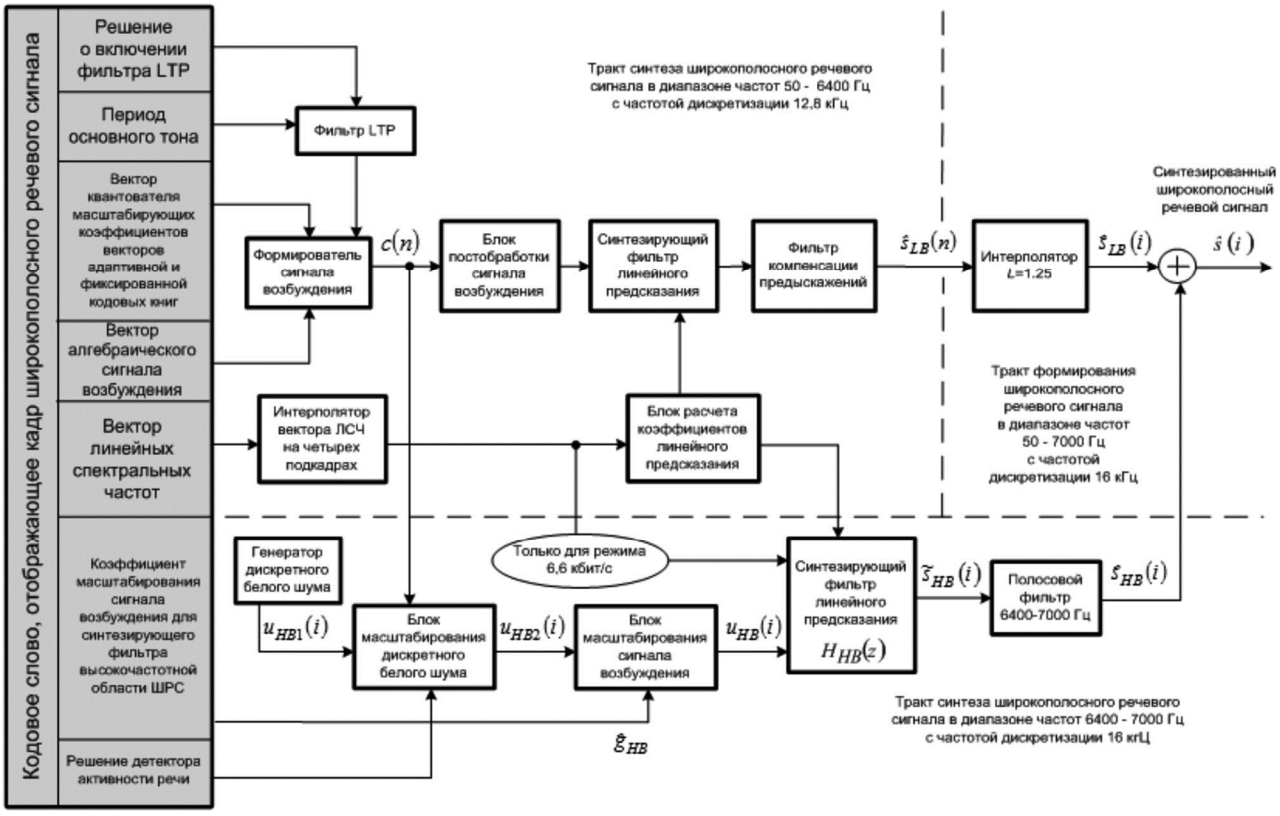

Рис. 2. Структура кодового слова, отображающего кадр ШРС, и структурная схема декодера AMR-WB

Таблица 1. Распределение бит кодового слова кодека AMR-WB по кодируемым параметрам текущего кадра ШРС для различных скоростей кодирования

Кодируемые параметры широкополосного речевого сигнала	Количество бит				
	1-й подкадр	2-й подкадр	3-й подкадр	4-й подкадр	Кадр ШРС
1	2	3	4	5	6
Для скорости кодирования 23,85 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	88	88	88	88	352
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Коэффициент масштабирования сигнала возбуждения для синтезирующего фильтра высокочастотной области ШРС	4	4	4	4	16
Общее количество бит	109	106	109	106	477
Для скорости кодирования 23,05 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	88	88	88	88	352
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	105	102	105	102	461
Для скорости кодирования 19,85 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4

Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	72	72	72	72	288
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	89	86	89	86	397
Для скорости кодирования 18,25 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	64	64	64	64	256
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	81	78	81	78	365
Для скорости кодирования 15,85 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	52	52	52	52	208
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	69	66	69	66	317
Для скорости кодирования 14,25 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	44	44	44	44	176
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	61	58	61	58	285
Для скорости кодирования 12,65 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	1	1	1	1	4
Период основного тона	9	6	9	6	30
Вектор алгебраического сигнала возбуждения	36	36	36	36	144
Вектор квантователя масштабирующих коэффициентов векторов адаптивной и фиксированной кодовых книг	7	7	7	7	28
Общее количество бит	53	50	53	50	253
Для скорости кодирования 8,85 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	46
Решение о включении фильтра LTP	8	5	8	5	26
Период основного тона	20	20	20	20	80
Вектор алгебраического сигнала возбуждения	6	6	6	6	24
Общее количество бит	34	31	34	31	177
Для скорости кодирования 6,6 кбит/с					
Решение детектора активности речи	–	–	–	–	1
Вектор линейных спектральных частот	–	–	–	–	36
Решение о включении фильтра LTP	8	5	5	5	26
Период основного тона	12	12	12	12	48
Вектор алгебраического сигнала возбуждения	6	6	6	6	24
Общее количество бит	26	23	23	23	132

Анализ широкополосного речевого сигнала в диапазоне частот 50-6400 Гц

Этап предварительной обработки сигнала

На этапе предварительной обработки текущий кадр исходного цифрового ШРС, представленный в формате 14-битной ИКМ, преобразовывается в цифровой сигнал с диапазоном частот 50-6400 Гц и частотой дискретизации 12,8 кГц последовательным выполнением следующих процедур:

– интерполяции с коэффициентом L = 4, повышающей частоту дискретизации до 64 кГц;

– низкочастотной фильтрации с частотой среза 6,4 КГц;

– децимации с коэффициентом M = 5, понижающей частоту дискретизации до 12, 8 кГц;

– высокочастотной фильтрации фильтром с частотой среза 50 Гц, обеспечиваемой передаточной функцией

$$H_h(z) = \frac{0{,}989502 - 1{,}979004 \cdot z^{-1} + 0{,}989502 \cdot z^{-2}}{1 - 1{,}978882 \cdot z^{-1} + 0{,}9799126 \cdot z^{-2}}.$$

Сформированный кадр $s_{\text{op}}(n),\ n = 0,1,2,...,255$ поступает на детектор активности речи и фильтр предыскажений.

Так как спектральная плотность мощности ШРС характеризуется бóльшей неравномерностью (наклоном), чем в узкополосной телефонии, для относительного выравнивания по мощности спектральных составляющих используется предыскажение обрабатываемого сигнала нерекурсивным цифровым фильтром первого порядка с передаточной функцией $H_{\text{pre-emph}}(z) = 1 - 0.68 \cdot z^{-1}$. Амплитудно-частотная характеристика фильтра представлена на рис. 3.

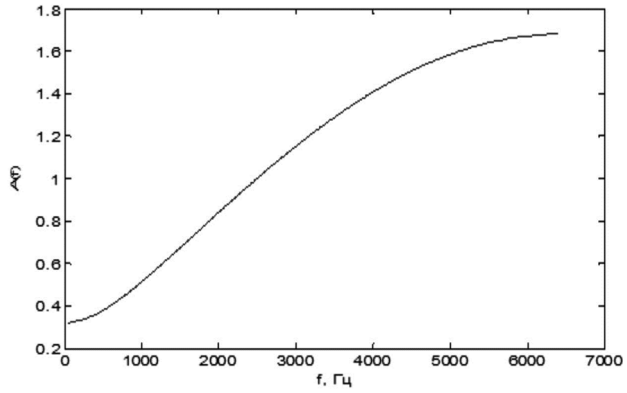

Рис. 3. Амплитудно-частотная функция фильтра предыскажения

Таким образом, в результате предварительной обработки формируется кадр речевого сигнала $s(n), n = 0,1,2,...,255$ с частотным диапазоном 50 – 6400 Гц, частотой дискретизации 12,8 кГц и относительно плоской спектральной плотностью мощности.

Этап идентификации фильтра кратковременного линейного предсказания

Параметрическая идентификация фильтра STP-анализа осуществляется однократно на длительности кадра и содержит последовательно выполняемые процедуры оконного взвешивания текущего кадра ШРС, расчета его автокорреляционной функции, расчёта коэффициентов линейного предсказания, определения и квантования линейных спектральных пар (линейных спектральных частот). Многовариантность представления параметров голосового тракта в этом случае объясняется удобством и меньшей вычислительной сложностью требуемых расчётных операций.

Формирование текущего кадра $s(n)$ осуществляется ассиметричной оконной функцией длительностью 30 мс, представляющей собой комбинацию из двух вариантов обобщенного окна Хэмминга (рис. 4):

$$w(n) = \begin{cases} 0{,}54 - 0{,}46\cos\left(\dfrac{2\pi n}{2L_1 - 1}\right);\ n = 0,...,L_1 - 1; \quad L_1 = 256; \\ \cos\left(\dfrac{2\pi(n - L_1)}{4L_2 - 1}\right);\ n = L_1,..,L_1 + L_2 - 1; \quad L_2 = 128; \end{cases}$$

при этом обеспечивается перекрытие соседних кадров на 5 мс, как показано на рис. 5.

Для взвешенного кадра $s'(n), n = 0,1,2,...,383$ рассчитывается автокорреляционная функция $r(k) = \sum_{n=k}^{383} s'(n) \cdot s'(n-k),\ k = 0, 1, 2, .., 16.$, по которой алгоритмом Левинсона-Дарбина определяются коэффициенты линейного предсказания (КЛП) 16-го порядка $a_m,\ m = 1, .., 16.$ Использование для расчёта КЛП предыскаженного речевого кадра позволяет в дальнейшем получить требуемую СПМ сигнала ошибки квантования.

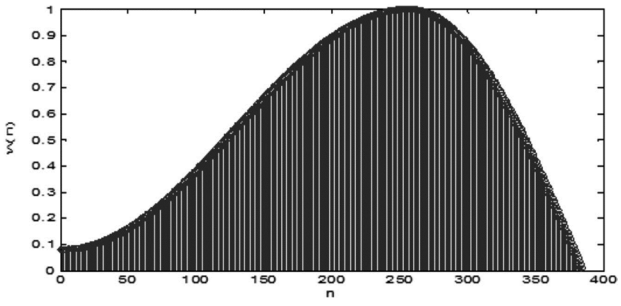

Рис. 4. Оконная функция взвешивания текущего кадра ШРС

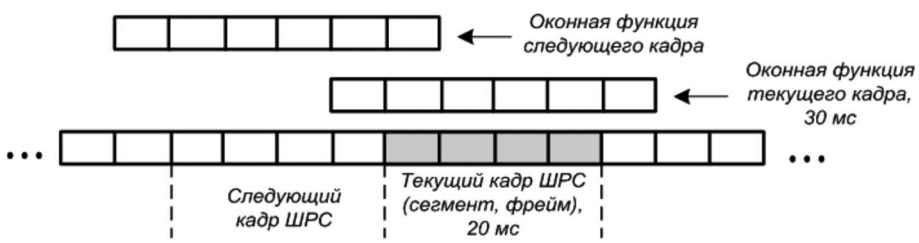

Рис. 5. Принцип оконного взвешивания кадров ШРС с перекрытием

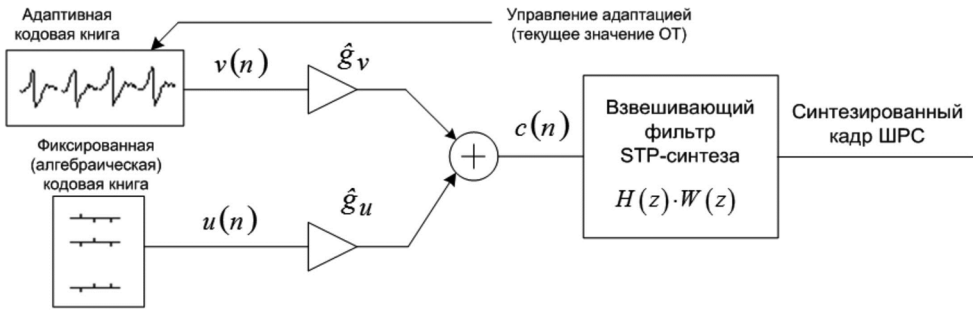

Рис. 6. Модель формирования сигнала возбуждения фильтра STP-синтеза

Конвертация КЛП в линейные спектральные пары (ЛСП) реализована на основе полиномов Чебышева.

В дальнейших процедурах кодирования ШРС используются как неквантованные, так и квантованные линейные спектральные частоты, для их квантования используется комбинация расщеплённого векторного квантования (SVQ – split vector quantization) и каскадного векторного квантования (MSVQ – multistage vector quantization). При этом неквантованные ЛСЧ определяют передаточную функцию анализирующего фильтра линейного предсказания $A(z)$, а квантованные ЛСЧ – её квантованную версию $\hat{A}(z)$. Существенно, что рассчитанные неквантованные и квантованные ЛСЧ в дальнейшей обработке используются только для четвёртого подкадра текущего кадра, а в первом – третьем подкадрах используются значения, полученные линейной интерполяцией ЛСЧ смежных кадров.

Этап формирования сигнала возбуждения для синтезирующего фильтра линейного предсказания

Модель сигнала возбуждения синтезирующего фильтра

Значимость оптимального выбора сигнала возбуждения для требуемого качества синтезированной речи подтверждается долей информационных ресурсов (размерности кодового слова), выделяемых для его двоичного представления: от 71,97 % при скорости кодирования 6,6 кбит/с до 89,8 % при скорости кодирования 23,05 кбит/с (табл. 1). Формирование сигнала возбуждения для различных скоростей кодирования имеет свои особенности, ниже представлены лишь наиболее общие процедуры.

Используемая модель формирования сигнала возбуждения для текущего подкадра ШРС определяется выражением $c(n) = \hat{g}_v \cdot v(n) + \hat{g}_u \cdot u(n), \quad n = 0, \ldots, 63,$ где $\hat{g}_v$ и $\hat{g}_u$ – квантованные масштабирующие коэффициенты, и представлена на рис. 6.

Непериодическая составляющая $\hat{g}_u \cdot u(n)$ сигнала возбуждения формируется масштабированием сигналов (векторов, кодовых слов) $u(n)$, содержащихся в фиксированной кодовой книге, построенной по алгебраическому принципу.

Периодическая (тоновая) составляющая $\hat{g}_v \cdot v(n)$ сигнала возбуждения формируется масштабированием векторов $v(n)$ адаптивной кодовой книги, которые определяются для каждого подкадра анализа с учётом текущего значения периода основного тона.

Определение периода основного тона

Определение периода ОТ, соответствующего текущему кадру ШРС, осуществляется в два этапа:

на первом этапе определяется «грубое» значение периода ОТ методом «разомкнутой петли»;

на втором этапе методом «замкнутой петли» уточняется ранее полученное «грубое» значение.

Анализ методом «разомкнутой петли» осуществляется на длительности 20 мс (при скорости кодирования 6,6 кбит/с) или 10 мс (в других скоростных режимах) на основе анализа кадра ШРС $s_w(n)$, формируемого фильтром перцептуального взвешивания. Процедура перцептуального взвешивания традиционно нацелена на перенос шумов квантования речевого сигнала в формантные области, что обеспечивает маскировку указанных шумов и улучшает слуховое восприятие речи. В рассматриваемом кодере с учётом значительного наклона СПМ широкополосного речевого сигнала процедура перцептуального взвешивания осуществляется последовательным выполнением следующих операций:

– относительным выравниванием СПМ речевого сигнала фильтром предыскажения $H_{pre-emph}(z)$;

– расчётом коэффициентов линейного предсказания по предыскажённому ШРС;

– использованием модифицированного фильтра перцептуального восприятия с передаточной функцией

$$W(z) = \frac{S_w(z)}{S(z)} = A(z/\gamma_1) \cdot H_{\text{de-emph}}(z),$$

где $A(z)$ – передаточная функция анализирующего фильтра линейного предсказания, сформированная на основе КЛП a_m, полученных конвертацией из неквантованных значений ЛСП, и отображающая формантную структуру текущего кадра;

– $\gamma_1 = 0{,}92$ – коэффициент перцептуального взвешивания;

– $H_{\text{de-emph}}(z) = H^{-1}_{pre-emph}(z) = \dfrac{1}{1 - 0{,}68 \cdot z^{-1}}$ – передаточная функция фильтра, устраняющего предыскажение исходного ШРС.

Тогда

$$W(z) = A(z/\gamma_1) \cdot H_{\text{de-emph}}(z) = 1 + \sum_{m=1}^{16} a_m \cdot \left(\frac{z}{0{,}92}\right)^{-m} \times$$

Рис. 7. Модель реализации двухэтапной процедуры анализа через синтез

$$\times \frac{1}{1-0{,}68\cdot z^{-1}}.$$

Таким образом, на выходе фильтра перцептуального взвешивания формируется сигнал

$$s_w(n) = s(n) + \sum_{m=1}^{16} a_m \cdot 0{,}92^m \cdot s(n-m) + 0{,}68 \cdot s_w(n-1),$$

$$n = 0,\ldots,L-1,$$

где длительность L соответствует 256 или 128 отсчетам.

Для уменьшения вычислительной сложности расчета периода ОТ предварительно используется децимация сигнала $s_w(n)$ с коэффициентом $L = 2$, при этом в дециматоре используется КИХ-фильтр 4-го порядка. Выходной сигнал $s_{wd}(n)$ дециматора содержит 128 или 64 отсчета на длительности кадра.

На заключительном этапе метода «разомкнутой петли» на основе анализа взвешенной автокорреляционной функции $C(d)$ сигнала $s_{wd}(n)$ в диапазоне от 17Т до 115Т (что соответствует диапазону частот основного тона примерно от 56 Гц до 377 Гц) определяется «грубое» значение T_{op} периода ОТ и коэффициент масштабирования g, соответствующие процедуре долговременного линейного предсказания (LTP).

Процедура анализа через синтез

Для определения оптимальных сигналов (векторов) возбуждения в адаптивной и фиксированной кодовых книгах (КК) используется процедура анализа через синтез, в которой в качестве синтезирующего фильтра применяется взвешивающий фильтр STP-синтеза с передаточной функцией

$$H(z)\cdot W(z) = A(z/\gamma_1)\cdot H_{\text{de-emph}}(z)/\hat{A}(z),$$

где $H(z) = 1/\hat{A}(z)$ – передаточная функция фильтра STP-синтеза, определенная по квантованным КЛП (ЛСЧ). Импульсная характеристика $h(n)$ взвешивающего фильтра STP-синтеза рассчитывается для каждого подкадра ШРС фильтрацией вектора коэффициентов КИХ-фильтра с передаточной функцией $A(z/\gamma_1)$, дополненного нулями, последовательно соединенными фильтрами с передаточными функциями $1/\hat{A}(z)$ и $H_{de-emph}(z)$, в дальнейшем она используется в процедурах определения оптимальных сигналов возбуждения адаптивной и фиксированной кодовых книг. Схематично реализация двухэтапной процедуры анализа через синтез представлена на рис. 7.

Критерием определения оптимальных сигналов возбуждения из адаптивной и фиксированной кодовых книг является минимум среднеквадратической взвешенной ошибки между оригинальной и синтезированной речью. Для каждого этапа реализации процедуры анализа через синтез для этого формируются целевые векторы, характеризующие оригинальный ШРС.

Первый этап анализа через синтез предназначен для определения оптимального вектора адаптивной кодовой книги и обеспечивается переводом переключателей П1 и П2 в положения 1. Целевой вектор $x(n)$, используемый на этом этапе, формируется реакцией взвешивающего фильтра STP-синтеза на сигнал ошибки линейного предсказания текущего подкадра $r(n) = s(n) - \sum_{m=1}^{16} \hat{a}_m s(n-m),\ n = 0,\ldots,63,$ рассчитанный по квантованным КЛП (рис. 8).

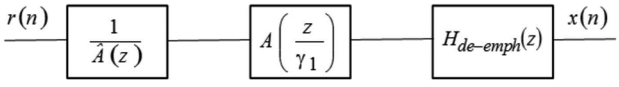

Рис. 8. Формирование целевого вектора $x(n)$

Первоначально в блоке определения оптимальных значений основного тона и масштабирующего коэффициента методом «замкнутой петли» осуществляется дальнейшее уточнение значения периода T_{op} в диапазоне от 34 до 231 периодов частоты 12800 Гц на основе использования целочисленных ($T_{op} \pm 7$) и дробных значений. Адаптация сигналов возбуждения $v(n)$ адаптивной КК реализуется фильтрацией исходных векторов на основе уточненного значения периода ОТ. Существенно, что при обработке ШРС целесообразен частотно-зависимый характер процедуры LTP, с этой целью на завершающем этапе формирования сигналов возбуждения в скоростных режимах 6,6 кбит/с и 8,85 кбит/с используется нерекурсивный ФНЧ с передаточной функцией $B_{LTP}(z) = 0{,}18\cdot z + 0{,}64 + 0{,}18\cdot z^{-1}$. При других скоростных режимах критерием принятия решения об использовании или неиспользовании данного фильтра является обеспечение минимальной мощности целевого вектора $x_2(n)$, принятое решение отображается одним битом кодового слова на каждом подкадре анализа, что соответствует 4 битам кодового слова текущего кадра ШРС (табл. 1).

Определение оптимального коэффициента масштабирования сигнала возбуждения адаптивной кодовой книги осуществляется в процессе анализа через синтез в диапазоне значений $0 \le g_v \le 1{,}2$.

Второй этап процедуры анализа через синтез предназначен для поиска оптимального вектора фиксированной кодовой книги и реализуется переводом переключателей П1 и П2 в положения 2. На этапе используется целевой вектор

$$x_2(n) = x(n) - g_v \cdot y(n), \qquad n = 0,...,63;$$

где $y(n) = v(n)*h(n)$ – реакция взвешивающего фильтра STP-синтеза на выбранный сигнал возбуждения адаптивной кодовой книги; g_v – неквантованный коэффициент масштабирования вектора адаптивной кодовой книги. Таким образом, из целевого вектора $x(n)$ вычитается реакция взвешивающего фильтра STP-синтеза, обеспечиваемая выбранным вектором сигнала возбуждения адаптивной КК.

Фиксированная кодовая книга относится к классу алгебраических, в различных скоростных режимах имеет различный объём алфавита (табл. 1) и характеризуется следующими особенностями формирования векторов $u(n)$:

– 64-мерные векторы разделяются на 2 (для скорости кодирования 6,6 кбит/с) или 4 (в остальных скоростных режимах) трека (сектора, дорожки) размерностью 32 или 16 отсчетов соответственно;

– в каждом треке, в основном состоящем из нулевых отсчётов, назначается фиксированное количество отсчётов, равных ± 1;

– позиции и значения ненулевых отсчётов любого k-го вектора $u_k(n)$ определяются особым («алгебраическим») способом по его номеру k и отображаются на определённых битовых позициях итогового кодового слова; такой способ формирования векторов $u(n)$ требует значительно меньшего объёма памяти в отличие от стохастических кодовых книг, в которых кодовые векторы хранятся в табличном виде.

В табл. 2 представлен пример формирования фиксированной кодовой книги для скоростного режима 12, 65 кбит/с. В этом режиме вклад фиксированной кодовой книги в кодовое слово текущего кадра ШРС (табл. 1) составляет 144 бита (по 36 бит для каждого подкадра: для каждого ненулевого отсчёта – 4 бита на номер позиции и 1 бит на значение первого ненулевого отсчёта в каждом треке).

Одновременно с формированием вектора $u(n)$ определяется оптимальный коэффициент его масштабирования g_u, обуславливающий энергию непериодического компонента $g_u \cdot u(n)$ сигнала возбуждения фильтра STP- синтеза.

Векторное квантование коэффициентов масштабирования

Рассчитанные коэффициенты масштабирования g_v и g_u подвергаются процедуре двумерного векторного квантования и отображаются в кодовом слове текущего кадра анализа 24 битами (для режимов 6,6 кбит/с и 8,85 кбит/с) или 28 битами (для остальных режимов). Квантованный сигнал возбуждения синтезирующего фильтра, как показано на рисунке 6, имеет вид

$$c(n) = \hat{g}_v \cdot v(n) + \hat{g}_u \cdot u(n), \quad n = 0,1,2,...,63.$$

Этап обновления состояния взвешивающего фильтра STP-синтеза

Расчёт целевого сигнала $x(n)$ для текущего подкадра анализа требует предварительного обновления начального состояния взвешивающего фильтра STP-синтеза. Обновление может быть достигнуто фильтрацией взвешенным фильтром STP-синтеза сигнала, представляющего разность между сигналом ошибки предсказания $r(n)$ и сигналом возбуждения $c(n)$, ,сформированным по результатам анализа предыдущего подкадра, с сохранением финального состояния фильтров. Такой подход требует тройного последовательного осуществления процедуры цифровой фильтрации фильтрами с передаточными функциями $1/\hat{A}(z)$, $A(z/\gamma)$ и $H_{de-emph}(z)$.

На практике используется упрощённая процедура обновления, требующая однократного применения процедуры фильтрации:

– фильтрацией фильтром с передаточной функцией $1/\hat{A}(z)$ при сигнале возбуждения $c(n)$, полученном по результатам анализа предыдущего подкадра, формируется кадр синтезированной речи $\hat{s}(n)$;

– очевидно, что разностный сигнал $e(n) = s(n) - \hat{s}(n)$ эквивалентен требуемой разности $r(n) - c(n)$, при этом начальное состояние фильтра $1/\hat{A}(z)$ определяется значениями сигнала $e(n), n = 48,49,...,63$;

Таблица 2. Формирование векторов $u(n), n = 0,1,2,...,63$ фиксированной кодовой книги для скоростного режима 12,65 кбит/с

Номер трека	Номера ненулевых отсчётов	Потенциальные позиции n ненулевых отсчётов в треках вектора $u(n)$
1	i_0, i_4	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

– обновление состояния фильтра $A(z/\gamma)\cdot H_{de-emph}(z)$ может быть достигнуто фильтрацией сигнала $e(n)$ с формированием сигнала взвешенной ошибки $e_w(n)$; однако указанный сигнал может быть получен расчетным путём:

$$e_w(n) = x(n) - \hat{g}_v \cdot y(n) = \hat{g}_u z(n)$$

где $z(n)$ = $u(n)$*$h(n)$ – реакция взвешивающего фильтра STP-синтеза на выбранный сигнал возбуждения фиксированной кодовой книги; при этом начальное состояние фильтра $A(z/\gamma)\cdot H_{de-emph}(z)$ определяется значениями сигнала $e_w(n), n = 48,49,...,63$.

Анализ широкополосного речевого сигнала в диапазоне частот 6400 –7000 Гц

Тракт анализа речевого сигнала с диапазоном частот от 6400 Гц до 7000 Гц используется только в скоростном режиме 23,85 кбит/с и формируется цифровым полосовым фильтром с конечной импульсной характеристикой, работающим на частоте дискретизации 16 кбит/с. Сигнал $s_{HB}(i)$ с выхода фильтра поступает на блок расчёта коэффициента масштабирования g_{HB}, назначение которого обусловлено используемым в декодере AMR-WB способом синтеза высокочастотной части ШРС (рис. 2), предусматривающим последовательное выполнение следующих процедур:

– генерирование дискретного сигнала белого шума $u_{HB1}(i),\ i = 0,1,...,79$ с частотой дискретизации 16 кГц;

– формирование шумового сигнала $u_{HB2}(i)$ на основе выравнивания мощностей сигнала $u_{HB1}(i)$ и сигнала возбуждения $c(n)$, используемого для синтезирующего фильтра в диапазоне частот 50-6400 Гц:

$$u_{HB1}(i)\cdot\sqrt{\sum_{n=0}^{63} c^2(n) \Big/ \sum_{i=0}^{63} {u_{HB1}}^2(i)};$$

– формирование сигнала возбуждения $u_{HB}(i)$ масштабированием сигнала $u_{HB2}(i)$ квантованным коэффициентом $\hat{g}_{HB}$, неквантованная версия которого рассчитана в кодере:

$$u_{HB}(i) = \hat{g}_{HB}\cdot u_{HB2(i)};$$

– синтез подкадра широкополосного дискретного сигнала $\tilde{s}_{HB}(i),\ i = 0,1,...,79$ на основе синтезирующего фильтра

$$H_{HB}(z) = \frac{1}{\hat{A}_{HB}(z)}$$

с последующим выделением цифровым полосовым фильтром высокочастотной составляющей $\hat{s}_{HB}(i),\ i = 0,1,...,79$ ШРС с диапазоном частот 6400 – 7000 Гц.

Передаточная функция $\hat{A}_{HB}(z)$ рассчитывается на основе рассчитанных ранее квантованных КЛП (ЛСЧ) с пересчетом частоты дискретизации с 12,8 кГц на 16 кГц:

$\hat{A}_{HB}\,z = \hat{A}(z/0{,}8)$.

Подобное преобразование обеспечивает пересчет частотной характеристики:

$$H_{16}(f) = H_{12.8}\left(\frac{12.8}{16}\cdot f\right),$$

что соответствует отображению частотной области 5,12 – 5,6 кГц фильтра $\hat{A}(z)$ на частотную область 6,4 – 7,0 кГц фильтра $\hat{A}_{HB}(z)$.

Таким образом, очевидно, что реализация в декодере процедуры синтеза речевого сигнала в диапазоне частот 6,4 – 7 кГц требует предварительного определения в кодере коэффициента масштабирования g_{HB}:

$$g_{HB} = \sum_{i=0}^{63}\left(s_{HB}(i)\right)^2 \Big/ \sum_{i=0}^{63}\left(s_{HB2}(i)\right)^2,$$

где $s_{HB2}(i)$ – речевой сигнал в полосе частот 6400 – 7000 Гц, синтезированный фильтром с передаточной функцией $H_{HB}(z)$ при возбуждении его сигналом $u_{HB2}(i)$. Квантованное значение $\hat{g}_{HB}$ коэффициента масштабирования, рассчитанное для текущего подкадра в скоростном режиме 23,85 кбит/с, отображается 4-мя битами кодового слова.

Для иных скоростей кодирования ШРС коэффициент g_{HB} кодером не определяется, его значение в декодере рассчитывается автономно по параметрам сегментов активной речи.

Заключение

Анализ процедур, применяемых в представленном кодеке, позволяет сформировать оценку уровня технологий, применяемых для адаптивного кодирования широкополосного речевого сигнала с переменной скоростью в задаче телефонии. Однако следует иметь в виду, что процедуры аналого-цифрового преобразования ШРС продолжают совершенствоваться, а в практику широкополосной телефонии внедряются новые алгоритмы обработки речи. С этой точки зрения значительный интерес для специалистов представляет широкополосный кодек для улучшенного речевого сервиса (EVS – Enhanced Voice Services), разработанный группой исследователей в рамках проекта 3GPP (3rd Generation Partnership Project) [5].

Литература

1. Потапова Р.К. Речь: коммуникация, информатика, кибернетика: Учеб. Пособие для вузов. М.: Радио и связь. 1997. 528 с.

2. Фланаган Д.Л. Анализ, синтез и восприятие речи. Перевод с англ. под редакцией А.А. Пирогова. М.: Связь. 1968.396 с.

3. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. 1995. Москва: ИПК издательство стандартов, 1996. 230 с.

4. ITU-T Recommendation G.722.2. Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB). 2003.

5. 3GPP TS 26.445. Codec for Enhanced Voice Services; Detailed Algorithmic Description (Release 15). 2018.

6. J.-P. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, «Fast CELP coding based on algebraic codes», in Acoustics, Speech, and Signal Processing, IEEE Int Conf (ICASSP'87), April 1987, pp. 1957-1960.