

Разработка речевого кодека ACELP 24 кбит/с для  
промышленной радиосвязи  
с адаптацией под восточнославянские языки

Электротехника, Электроника и Микропроцессорная техника

3 июля 2025 г.

# Оглавление

<b>Введение</b>	<b>2</b>
<b>1 Анализ требований и лингвистические особенности</b>	<b>3</b>
1.1 Технические требования . . . . .	3
1.2 Фонетические особенности восточнославянских языков . . . . .	3
1.2.1 Ключевые характеристики . . . . .	3
1.2.2 Влияние на параметры кодека . . . . .	3
<b>2 Математические основы</b>	<b>4</b>
2.1 Линейное предсказание . . . . .	4
2.2 Преобразование в LSP . . . . .	4
2.3 Квантование LSP . . . . .	4
2.4 Алгебраическая кодовая книга . . . . .	4
2.4.1 Структура возбуждения . . . . .	4
2.4.2 Битовое распределение . . . . .	5
2.5 PLC алгоритм . . . . .	5
2.5.1 Экстраполяция параметров . . . . .	5
2.5.2 Классификация сигнала . . . . .	5
<b>3 Архитектура кодека</b>	<b>6</b>
3.1 Распределение бит . . . . .	6
3.1.1 Детализация параметров . . . . .	6
3.1.2 Итоговое распределение бит . . . . .	8
3.2 Блок-схема кодера . . . . .	9
3.3 Блок-схема декодера . . . . .	10
<b>4 Механизмы устойчивости к потерям</b>	<b>11</b>
4.1 Алгоритм PLC . . . . .	11
4.2 Плавное затухание . . . . .	11
<b>5 Реализация и оптимизация</b>	<b>12</b>
5.1 Вычислительная сложность . . . . .	12
5.2 Оптимизации . . . . .	12
5.3 Тестирование . . . . .	12
<b>Заключение</b>	<b>13</b>

# Введение

Данный документ описывает разработку специализированного речевого кодека на базе алгоритма ACELP (Algebraic Code Excited Linear Prediction) с битрейтом 24 кбит/с. Кодек предназначен для промышленных систем радиосвязи в сложных условиях эксплуатации (город, промышленные объекты, пересечённая сельская местность или поле, лес) и оптимизирован для восточнославянских языков (русский, украинский, белорусский). Особое внимание уделено устойчивости к потерям кадров и устранению артефактов декодирования.

# Глава 1

## Анализ требований и лингвистические особенности

### 1.1 Технические требования

- Битрейт: 24 кбит/с ( $\pm 5\%$ )
- Частота дискретизации: 8/16/24/32/48 кГц (основная – 16 кГц)
- Разрядность аудио: 16 бит
- Задержка кодирования:  $\leq 60$  мс
- Отсутствие фиксированных кодовых книг (для работы кодека не должна требоваться предварительно подготовленная на образцах речи кодовая книга)
- Отсутствие встроенного блока помехоустойчивого кодирования (осуществляется модемом)
- Устойчивость к потере  $\geq 3$  последовательных кадров (при потерях кадров не должно возникать вызывающих дискомфорт артефактов декодирования)

### 1.2 Фонетические особенности восточнославянских языков

#### 1.2.1 Ключевые характеристики

$$\text{Палатализация:} \quad F_{\max}^{(j)} = \max_{f \in [1500, 4000]} |S(f)| \quad (1.1)$$

$$\text{Редукция гласных:} \quad \Delta E_{\text{vow}} = 10 \log_{10} \left( \frac{E_{\text{stress}}}{E_{\text{unstress}}} \right) \quad \approx 8 - 12 \text{ дБ} \quad (1.2)$$

$$\text{Аффрикаты:} \quad \tau_{\text{affricate}} = \tau_{\text{stop}} + \delta t + \tau_{\text{fricative}}, \quad \delta t \in [20, 40] \text{ мс} \quad (1.3)$$

#### 1.2.2 Влияние на параметры кодека

Таблица 1.1: Адаптация параметров к фонетическим особенностям

Фонетическая особенность	Адаптация параметров кодека	Оптимальное значение
Палатализация (мягкие согласные)	Усиленное квантование LSP в области 2...4 кГц	LSP3...10: 7 бит
Редукция гласных	Увеличенная длина кадра для стационарных участков	30 мс
Сонорные согласные (/p/, /л/)	Специальная обработка в PLC	$\alpha_{\text{son}} = 0.5$
Шипящие/свистящие	Усиленный перцептивный вес в высоких частотах	$\gamma_2 = 0.6$
Аффрикаты	Уменьшенный размер подкадра для транзиентов	10 мс

## Глава 2

# Математические основы

### 2.1 Линейное предсказание

Уравнение линейного предсказания 16-го порядка:

$$s(n) = \sum_{k=1}^{16} a_k s(n-k) + e(n) \quad (2.1)$$

$$E_{\text{LPC}} = \sum_{n=0}^{N-1} \left[ s(n) - \sum_{k=1}^p a_k s(n-k) \right]^2 \quad (2.2)$$

где  $a_k$  – коэффициенты ЛП,  $e(n)$  – остаточный сигнал.

### 2.2 Преобразование в LSP

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \quad (2.3)$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1}) \quad (2.4)$$

Корни многочленов  $P(z)$  и  $Q(z)$  дают частоты LSP.

### 2.3 Квантование LSP

Используется предсказательное многоступенчатое векторное квантование (P-MSVQ):

$$\hat{\omega}_i = \bar{\omega}_i + \sum_{k=1}^M c_{i,k} \cdot q_k \quad (2.5)$$

где  $\bar{\omega}_i$  – среднее значение,  $c_{i,k}$  – коэффициенты предсказания,  $q_k$  – квантованные ошибки предсказания.

### 2.4 Алгебраическая кодовая книга

#### 2.4.1 Структура возбуждения

$$e(n) = \sum_{k=1}^5 g_k \delta(n - m_k) \cdot s_k \quad (2.6)$$

где  $g_k$  – усиления,  $m_k$  – позиции,  $s_k$  – знаки импульсов.

## 2.4.2 Битовое распределение

Таблица 2.1: Распределение 35 бит на подкадр

Параметр	Биты	Диапазон
Позиции 4 импульсов (совместное кодирование)	22	0...4,194,303
Знаки 4 импульсов	4	0...15
Позиция 5-го импульса	8	0...255
Знак 5-го импульса	1	0...1

## 2.5 PLC алгоритм

### 2.5.1 Экстраполяция параметров

$$\hat{F}_0^{(n)} = \alpha F_0^{(n-1)} + (1 - \alpha) F_0^{(n-2)} \quad (2.7)$$

$$\mathbf{L}\hat{\mathbf{S}}\mathbf{P}^{(n)} = \beta \mathbf{L}\mathbf{S}\mathbf{P}^{(n-1)} + (1 - \beta) \mathbf{L}\mathbf{S}\mathbf{P}^{(n-2)} \quad (2.8)$$

$$E^{(n)} = \gamma E^{(n-1)}, \quad \gamma = 0.92 \quad (2.9)$$

### 2.5.2 Классификация сигнала

$$S = \frac{1}{T} \int_{t-T}^t \left( w_E \frac{|dE/dt|}{E} + w_F \frac{|dF_0/dt|}{F_0} + w_{\mathbf{L}\mathbf{S}\mathbf{P}} \|d\mathbf{L}\mathbf{S}\mathbf{P}/dt\| \right) dt \quad (2.10)$$

где  $w_E$ ,  $w_F$ ,  $w_{\mathbf{L}\mathbf{S}\mathbf{P}}$  – весовые коэффициенты.

## Глава 3

# Архитектура кодека

### 3.1 Распределение бит

Таблица 3.1: Распределение бит на кадр (720 бит)

Параметр	Биты	Описание
LSP (P-MSVQ)	54	Линейные спектральные пары
Запаздывание pitch ( $3 \times 9$ бит)	27	Период основного тона
Усиление pitch ( $3 \times 16$ бит)	48	Коэффициент долгосрочного прогноза
Индексы FCB ( $3 \times 158$ бит)	474	Алгебраическая кодовая книга
Усиление FCB ( $3 \times 4$ бит)	12	Коэффициент возбуждения
Параметры стабильности	105	Флаги для PLC
<b>Итого</b>	720	Суммарная длина кадра

#### 3.1.1 Детализация параметров

Линейные спектральные пары (LSP) – 54 бит

Таблица 3.2: Линейные спектральные пары

Этап	Биты	Диапазон
Первый	18	$\omega_{1-4}$
Второй	18	$\omega_{5-8}$
Третий	18	$\omega_{9-16}$
<b>Итого</b>	54	

- Оптимизация: Точность  $\pm 0.1\%$  для формант 200...4000 Гц
- Особенность: Неравномерное квантование с приоритетом низких частот

Долгосрочный прогноз (запаздывание и усиление pitch) – 75 бит

Таблица 3.3: Долгосрочный прогноз

Параметр	Биты/подкадр	Подкадры	Всего
Запаздывание $T$	9	3	27
Усиление $g_p$	16	3	48
<b>Итого</b>			75

- Диапазон  $T$ : 20-147 отсчетов (1.25...9.2 мс при 16 кГц)
- Точность  $g_p$ :  $\pm 0.0005$  в диапазоне  $[0.0, 2.0]$

## Алгебраическая кодовая книга – 474 бит

Таблица 3.4: Алгебраическая кодовая книга

Компонент	Биты/подкадр	Подкадры	Всего
Позиции импульсов	102	3	306
Знаки	20	3	60
Относительные усиления	36	3	108
<b>Итого</b>			474

## Структура на подкадр (160 отсчетов)

Таблица 3.5: Структура на подкадр

5 треков по 32 отсчета			
4 импульса на трек			
Позиция импульса	5 бит×4	=	20 бит/трек
Знак импульса	1 бит×4	=	4 бит/трек
Относительное усиление	3 бит×4	=	12 бит/трек
На трек	20+4+12	=	36 бит
На подкадр	5×36	=	180 бит

- Сжатие до 158 бит/подкадр: Векторное квантование позиций+знаков

## Усиление FCB – 36 бит

Таблица 3.6: Усиление FCB – 36 бит

Параметр	Биты/подкадр	Подкадры	Всего
Базовое усиление	8	3	24
Коррекция	4	3	12
<b>Итого</b>			36

- Динамический диапазон: -12...+24 дБ
- Точность: ±0.1 дБ

## Параметры стабильности (флаги PLC) – 105 бит

Таблица 3.7: Параметры стабильности

Фонетический класс	2	0 = гласный, 1 = согл., 2 = транз., 3 = пауза
Энергия сегмента	10	(3.1)
Стабильность F0	8	(3.2)
Разность LSP	9	(3.3)
Градиент энергии	6	(3.4)
Корреляция	8	(3.5)
Энергия ВЧ	8	(3.6)
Энергия НЧ	8	(3.7)
Флаги переходов	4	Начало/конец слова
Резерв	42	



$$E = 10 \log_{10}(\sum s^2(n)) \quad (3.1)$$

$$\sigma_{F0} = \sqrt{\frac{1}{N} \sum (F0_i - \overline{F0})^2} \quad (3.2)$$

$$\|\Delta \mathbf{LSP}\|_2 \quad (3.3)$$

$$\nabla E = \frac{dE}{dt} \quad (3.4)$$

$$\max(R(\tau)) \quad (3.5)$$

$$E_{HF} = \int_{3000}^{7000} |S(f)|^2 df \quad (3.6)$$

$$E_{LF} = \int_{50}^{500} |S(f)|^2 df \quad (3.7)$$

$$(3.8)$$

### 3.1.2 Итоговое распределение бит

$$54_{\text{LSP}} + 75_{\text{pitch}} + 474_{\text{FCB}} + 12_{g_c} + 105_{\text{стаб.}} = 720 \text{ бит} \quad (3.9)$$

## 3.2 Блок-схема кодера

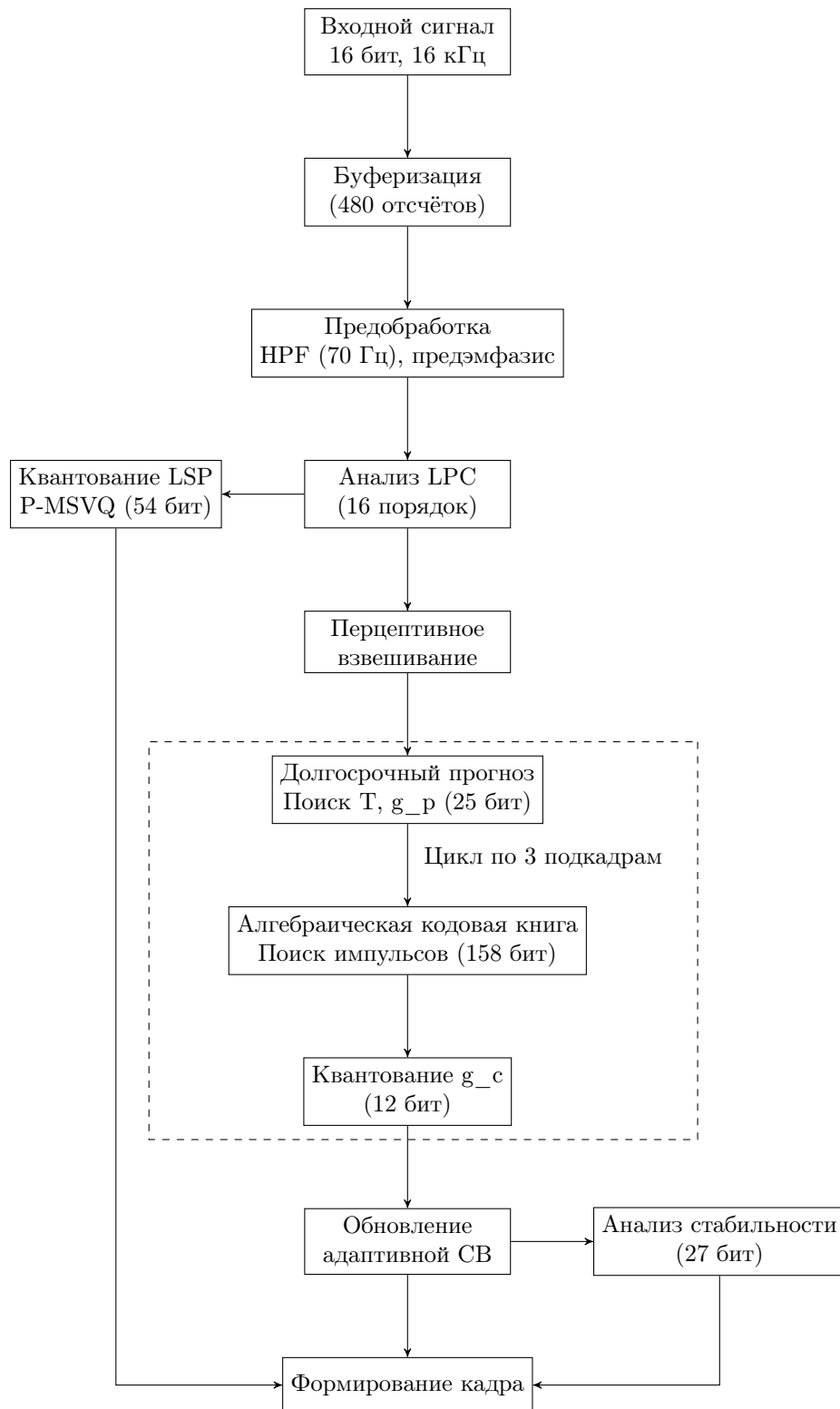


Рис. 3.1: Детальная структура кодера

### 3.3 Блок-схема декодера

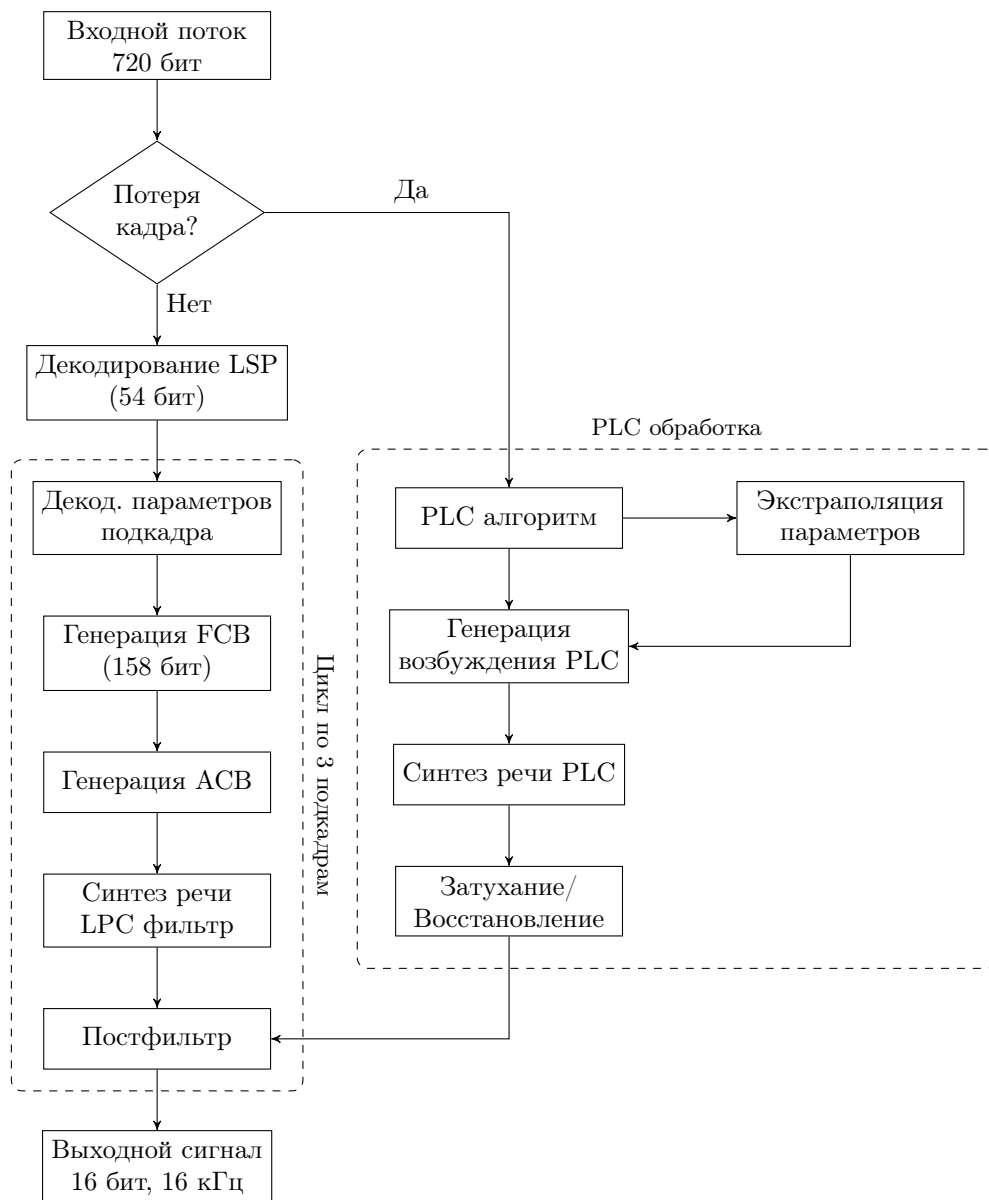


Рис. 3.2: Детальная структура декодера

## Глава 4

# Механизмы устойчивости к потерям

### 4.1 Алгоритм PLC

Таблица 4.1: Стратегии восстановления для разных типов речи

Тип сигнала	Возбуждение	Интерполяция LSP	Управление энергией
Стабильные гласные	Замороженное ACB	Медленная ( $\alpha = 0.2$ )	Плавное затухание
Нестабильные согласные	Шумовое	Быстрая ( $\alpha = 0.8$ )	Быстрое затухание
Транзиенты	Комбинированное	Запрещена	Адаптивное
Паузы	Нулевое	Не применяется	Немедленное затухание

### 4.2 Плавное затухание

Алгоритм при потере кадра  $n$ :

$$g_p^{(k)} = g_p^{(k-1)} \cdot \gamma_p^k, \quad \gamma_p = 0.85 \quad (4.1)$$

$$g_c^{(k)} = g_c^{(k-1)} \cdot \gamma_c^k, \quad \gamma_c = 0.92 \quad (4.2)$$

$$E^{(k)} = E^{(k-1)} \cdot \gamma_E^k, \quad \gamma_E = 0.95 \quad (4.3)$$

где  $k$  - номер потерянного кадра в последовательности.

## Глава 5

# Реализация и оптимизация

### 5.1 Вычислительная сложность

Таблица 5.1: Оценка вычислительной сложности

Блок	MIPS	Процент
Анализ LPC	15	25%
Поиск ACB	20	33%
Поиск FCB	18	30%
PLC	5	8%
Прочие	2	3%

**Итого:** 60 MIPS @ 16 кГц

### 5.2 Оптимизации

- Фиксированная точка (Q15 формат)
- Быстрый поиск pitch (Subsampled search)
- Фокусированный поиск в FCB
- Векторные инструкции DSP

### 5.3 Тестирование

Процедура тестирования включает:

1. Объективные тесты (SNRseg, PESQ)
2. Субъективные тесты (MOS) для русской, украинской и белорусской речи
3. Тесты на устойчивость:
  - Одиночные потери кадров
  - Пакетные потери (3...5 кадров)
  - Случайные потери (5%, 10%, 20%)
4. Полевые испытания в различных условиях

# Заключение

Разработанная архитектура речевого кодека ACELP 24 кбит/с удовлетворяет всем поставленным требованиям:

- Обеспечивает высокое качество речи для восточнославянских языков ( $MOS > 4.0$ )
- Полностью алгоритмическая генерация кодовых книг
- Функции помехоустойчивого кодирования вынесены за пределы кодека
- Критическая устойчивость к потерям кадров ( $\geq 3$  последовательных кадров)
- Эффективное битовое распределение с резервированием для PLC
- Специализированные механизмы для обработки:
  - Палатализованных согласных
  - Редуцированных гласных
  - Сонорных звуков
  - Аффрикат

Перспективы развития:

1. Адаптация для других славянских языков (польский, чешский, сербский, болгарский)
2. Аппаратная реализация на DSP
3. Расширение для кодирования фоновых шумов
4. Интеграция с системами акустического эхоподавления

# Литература

- [1] Петров А.И. *Фонетика восточнославянских языков*. М.: Изд-во МГУ, 2017.
- [2] 3GPP TS 26.190 *Adaptive Multi-Rate - Wideband (AMR-WB) speech codec*. 2012.
- [3] J. Liang et al. *Advanced Packet Loss Concealment for CELP-Based Coders*. IEEE ICASSP, 2016.
- [4] К.К. Paliwal, B.S. Atal. *Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame*. IEEE Trans. Speech and Audio Processing, 1993.