

УДК 621.391.266

В.Ю. Семенов

МЕТОДЫ ВЫЧИСЛЕНИЯ И КОДИРОВАНИЯ ПАРАМЕТРОВ АВТОРЕГРЕССИОННОЙ МОДЕЛИ РЕЧИ ПРИ РАЗРАБОТКЕ ВОКОДЕРА НА БАЗЕ СИГНАЛЬНОГО ПРОЦЕССОРА С ФИКСИРОВАННОЙ ТОЧКОЙ

Ключевые слова: авторегрессионная модель, рекурсия Левинсона–Дарбина, линейные спектральные частоты, период основного тона, арифметика с фиксированной точкой.

Введение

Развитие современных приложений цифровой связи, таких как сотовая и IP-телефония, привело к необходимости разработки высокоэффективных алгоритмов кодирования речи, обеспечивающих большую степень сжатия речевой информации. Алгоритмы кодирования речи можно классифицировать на три основные группы в зависимости от степени сжатия: высокоскоростные (скорость кодирования от 16 до 64 кбит/с); среднескоростные (4–16 кбит/с); низкоскоростные (менее 4 кбит/с).

Методы среднескоростного и низкоскоростного сжатия, важность которых в настоящее время особенно высока, основаны на моделях речеобразования. Наибольшее распространение получила авторегрессионная (АР) модель речеобразования [1], в которой речевой сигнал моделируется как результат прохождения управляющего процесса через авторегрессионный фильтр. АР-модель может быть представлена в разностной форме:

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + gw(n), \quad (1)$$

где $s(n)$ — речевой сигнал; $w(n)$ — управляющий процесс; g — коэффициент усиления; $a_k, k = 1, \dots, p$, — АР-коэффициенты. Порядок АР-модели p выбирается, как правило, равным 10.

Управляющий процесс $w(n)$ моделирует поток воздуха на выходе голосовых связок человека. Его простейшая аппроксимация представляет собой белый шум в случае произнесения глухих звуков или последовательность импульсов, идущих один за другим с периодом колебания голосовых связок (периодом основного тона).

Основным свойством параметров АР-модели является то, что они остаются практически неизменными на отрезках длиной 10–30 мс (свойство квазистационарности речевых сигналов). Таким образом, задача кодирования речи разбива-

ется на две подзадачи: кодирование возбуждающего процесса и кодирование АР-параметров. Типовая схема алгоритма кодирования речи, основанного на АР-модели, отображена на рис. 1.

При построении вокодеров большое внимание должно уделяться аппаратной базе, на которой они реализуются. В частности, ключевую роль играет выбор типа арифметики сигнального процессора (с фиксированной или плавающей точкой). Процессоры с плавающей точкой (floating point) характеризуются более широким динамическим диапазоном переменных, отсутствием проблем с переполнениями и ошибками округления, относительной простотой программирования и, как следствие, более высокой коммерческой стоимостью. Напротив, процессоры с фиксированной точкой (fixed-point) характеризуются малым размером, низким энергопотреблением и меньшей коммерческой стоимостью.

Цель данной публикации — разработка методов вычисления и кодирования параметров низкоскоростных вокодеров (1200 и 2400 бит/с) на базе цифрового сигнального процессора с фиксированной точкой ADSP-2191. Построение таких вокодеров является актуальной научно-технической задачей, имеющей важное прикладное значение.

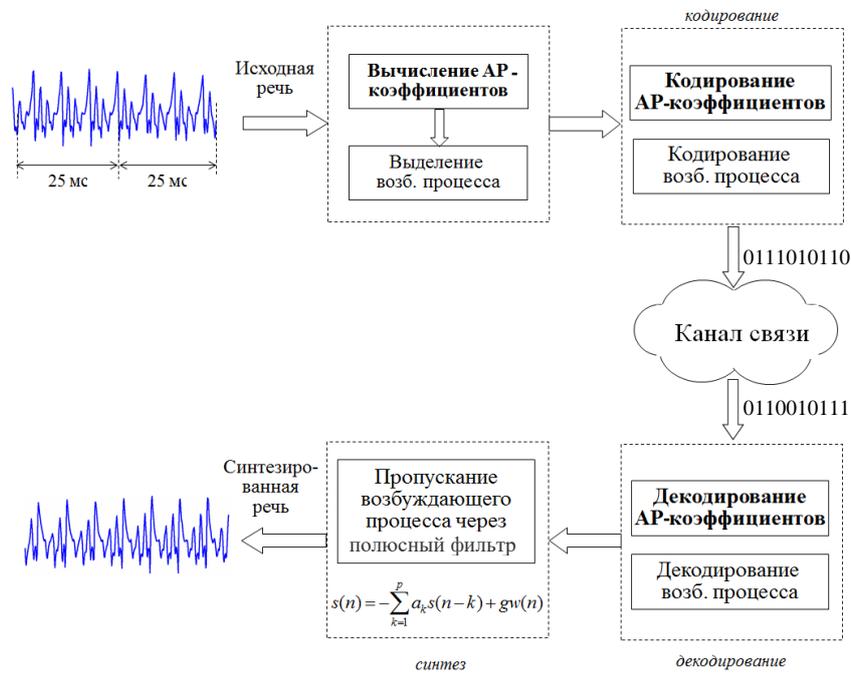


Рис. 1

Процедура анализа

Вычисление авторегрессионных коэффициентов. Традиционно АР-коэффициенты определяются с помощью методов линейного предсказания, в частности рекурсии Левинсона–Дарбина [1]:

$$\begin{cases} k_i = [r(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} r(i-j)] / e_{i-1}; \\ a_i^{(i)} = k_i; \\ a_j^{(i)} = a_j^{(i-1)} - k_i a_{j-j}^{(i-1)}, \quad 1 \leq j \leq i-1; \\ e_i = (1 - k_i^2) e_{i-1}, \quad 1 \leq i \leq p. \end{cases} \quad (2)$$

Действия в алгоритме (2) выполняются для $i=1, \dots, p$. В конце цикла присваивается $a_i = a_i^{(p)}$, $i=1, \dots, p$, что позволяет найти искомые АР-коэффициенты. Переменные e_i , $i=1, \dots, p$, играют роль ошибок линейного предсказания, а промежуточные переменные k_i , $i=1, \dots, p$, называются коэффициентами отражения и лежат в диапазоне $[-1, 1]$.

Однако переменные e_i , k_i , $a_j^{(i)}$ являются действительными, в то время как реализация на процессорах на выбранном сигнальном процессоре требует переход к переменным в формате с фиксированной точкой. С учетом специфики процессора ADSP 2191 выбран формат целочисленных переменных (16,0 и 32,0). Для этого предварительно были установлены динамические диапазоны переменных. Пример для коэффициента a_1 приведен на рис. 2. Примеры динамических диапазонов десяти АР-коэффициентов приведены в табл. 1. Здесь и далее тестовые эксперименты проводились на базе речевых сигналов, в формировании которых участвовали 16 мужчин и 11 женщин. Были представлены украинский и английский (США) языки. Общая их продолжительность составила около восьми минут для мужчин и около шести минут для женщин.

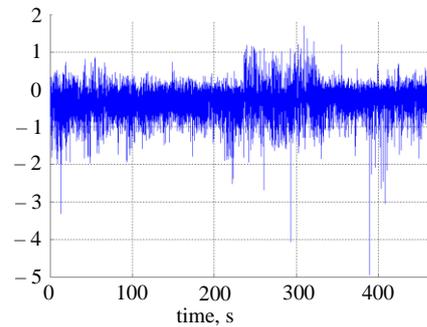


Рис. 2

Таблица 1

a	$\min\{a_i\}$	$\max\{a_i\}$
a_1	-3,16	1,51
a_2	-1,37	4,46
a_3	-4,43	1,34
a_4	-1,24	4,63
a_5	-4,95	1,68
a_6	-1,05	4,65
a_7	-4,44	0,96
a_8	-0,84	4,37
a_9	-3,00	0,90
a_{10}	-0,45	0,91

Таблица 2

A	$\min\{A_i\}$	$\max\{A_i\}$
A_1	-12970	6209
A_2	-5624	18290
A_3	-18185	5492
A_4	-5118	18984
A_5	-20283	6913
A_6	-4336	19086
A_7	-18205	3964
A_8	-3459	17910
A_9	-12307	3701
A_{10}	-1856	3749

Из табл. 1 следует, что для перехода в 16-разрядный целочисленный диапазон $[-32767, 32768]$ АР-коэффициенты должны быть умножены на коэффициент $M_a = 2^{12}$. Диапазоны значений целочисленных АР-коэффициентов A_i представлены в табл. 2. Соответственно коэффициенты отражения должны быть домножены на коэффициент $M_k = 2^{15}$.

Таким образом, можно сделать замену переменных:

$$\begin{cases} K_i = M_k \times k_i; \\ A_i = M_a \times a_i. \end{cases}$$

Затем перепишем алгоритм (2) в целых числах:

$$\begin{cases} K_i = [M_a R(i+1) - \sum_{j=1}^{i-1} A_j^{(i-1)} R(i-j)] M_k / [M_a E_{i-1}]; \\ A_i^{(i)} = M_a K_i / M_k; \\ A_j^{(i)} = [M_k A_j^{(i-1)} - K_i A_{i-j}^{(i-1)}] / M_k, 1 \leq j \leq i-1; \\ E_i = [M_k^2 - K_i^2] / E_{i-1} M_k^2, 1 \leq i \leq p. \end{cases} \quad (3)$$

Сравним вычислительные погрешности разработанного целочисленного алгоритма (3) с целочисленной рекурсией Шура, использовавшейся для вычисления АР-коэффициентов в программно-аппаратной реализации вокодера MELP [2]. Среднеквадратические погрешности вычисления коэффициентов представлены в табл. 3. Среднеквадратическая погрешность e_{LD} относится к реализации рекурсии Левинсона–Дарбина (3) и e_{Sh} — реализации рекурсии Шура.

Как видно из табл. 3, предложенный метод обладает существенно меньшими среднеквадратическими погрешностями целочисленного расчета АР-коэффициентов. Также проведено сравнение погрешности методов в рамках меры Итакуры [3] для предложенной реализации рекурсии Левинсона–Дарбина и рекурсии Шура (табл. 4). Как видно, предложенный метод существенно снижает погрешность вычисления коэффициентов:

$$d_I = \frac{1}{2\pi} \min_{\lambda > 0} \int_{-\pi}^{\pi} |\lambda F(\theta/G(\theta)) - \ln \lambda F(\theta/G(\theta)) - 1| d\theta.$$

Таблица 3

A	e_{LD}	e_{Sh}
A_1	25,2	34,3
A_2	38,2	51,4
A_3	42,4	56,6
A_4	46,3	64,5
A_5	50,5	100,0
A_6	48,6	87,1
A_7	46,6	123,2
A_8	39,0	145,0
A_9	36,2	85,7
A_{10}	23,3	230,6

Таблица 4

d	Предложенный метод	Рекурсия Шура
d_I по всем фреймам	0,0007	0,15
d_I по 95 % фреймов	0,0003	0,05
d_I по речевым фреймам	0,0008	0,17

Трансформация АР-коэффициентов в линейные спектральные частоты.

Непосредственное кодирование АР-коэффициентов неприменимо, поскольку приводит к неустойчивости синтезирующего фильтра приемника. Поэтому АР-параметры перед кодированием обычно преобразуются нелинейным образом в эквивалентный набор параметров, например в коэффициенты отражения или логарифмы отношений поперечных сечений модели речевого тракта [3]. Но наиболее популярным способом кодирования АР-параметров является их преобразование в набор линейных спектральных частот (ЛСЧ) или излучательных спектральных частот (см., например, обзор в [4]), квантование которых наиболее эффективно по сравнению с другими представлениями АР-модели. По определению (см., например, [4]) ЛСЧ являются аргументами комплексных корней полиномов:

$$\begin{cases} G_1(z) = \frac{A(z) + z^{-p-1}A(z^{-1})}{1 + z^{-1}}, \\ G_2(z) = \frac{A(z) - z^{-p-1}A(z^{-1})}{1 - z^{-1}}, \end{cases}$$

где $A(z) = 1 + \sum_{k=1}^p a_k z^{-k}$ — полином, состоящий из исходных АР-коэффициентов.

В [4] показано, что вычисление ЛСЧ сводится к поиску корней пары полиномиальных уравнений

$$\sum_{k=0}^M r_k^{(m)} x^{M-k} = 0 \quad (m = 1, 2), \quad (4)$$

где коэффициенты $r_k^{(m)}$ связаны с параметрами исходной АР-модели линейными соотношениями [4]. ЛСЧ связаны с корнями уравнений (4) обратным нелинейным преобразованием $\omega = \arccos(x)$.

Для нахождения ЛСЧ применен метод решения нелинейных уравнений [5], основная идея которого состоит в разбиении исходного интервала поиска корней на подинтервалы, в каждом из которых либо отсутствуют корни, либо выполняется критерий единственности корня. Для ускорения работы алгоритма также использовано свойство межфреймовой упорядоченности ЛСЧ [6].

Проведем сравнение между предложенным методом и методом Кабала–Рамачандрана (КР) [4], который широко используется в задачах обработки речевых сигналов. Поскольку метод КР использует уточнение корней методом бисекции, рассмотрена его ускоренная версия, основанная на уточнении корней методом Ньютона.

В табл. 5 представлено среднее, минимальное и максимальное число операций, требуемое для метода КР, его ускоренной версии и предложенного метода расчета ЛСЧ. В качестве критерия останова поиска корня взято условие $|x_k - x_{k-1}| < 10^{-3}$, где x_{k-1} , x_k — аппроксимации корня, полученные на смежных итерациях.

Таблица 5

Метод	Среднее число операций	Минимальное число операций	Максимальное число операций
Метод КР	2150	2150	2150
Ускоренный метод КР	1498	1491	1557
Предложенный метод	629	428	1428

Из табл. 5 следует, что предложенный метод обеспечивает сокращение вычислительных затрат по сравнению с методом КР в 3,42 раза. Сокращение числа операций по сравнению с ускоренным методом КР составляет 2,38 раз.

Важной особенностью разработанного метода является то, что его максимальное число операций меньше по сравнению с минимальным числом операций для ускоренного метода КР, что свидетельствует о его предпочтении для реализации в системах реального времени.

Вычисление периода основного тона. Существует большое многообразие методов для решения этой задачи [1]. Применен подход, ранее использованный

для идентификации периодических структур в дыхательных шумах [7]. В его основе лежит анализ автокорреляционной функции, которая для управляющего процесса $w(n)$ длиной T отсчетов имеет вид $R(t) = \sum_{k=0}^{T-t-1} w(k)w(k+t)$.

Известно, что если последовательность $w(n)$ имеет период t_0 , то ее автокорреляционная функция имеет пик в этой точке, а также побочные пики в кратных значениях, как это представлено на рис. 3.

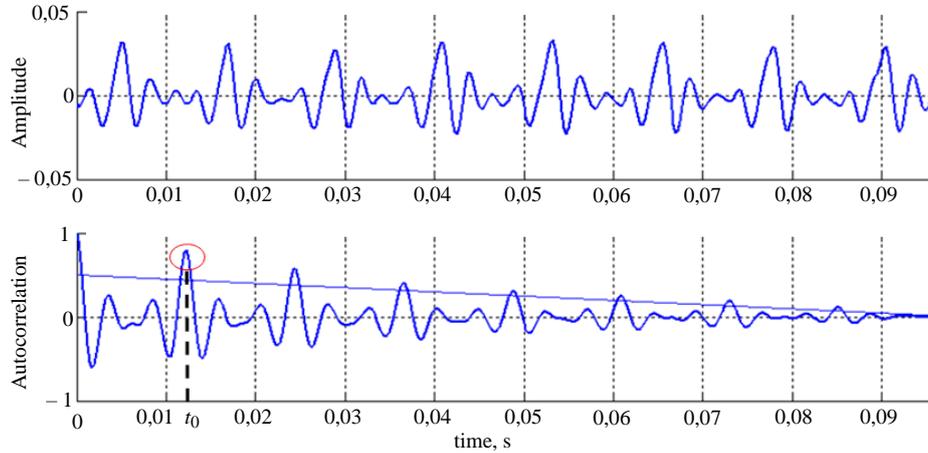


Рис. 3

Предполагая наличие в сигнале смеси периодической и шумовой компонент ($w(n) = w_p(n) + w_n(n)$), критерий преобладания в сигнале периодической компоненты можно представить в виде

$$E_p \geq E_n, \quad (5)$$

где E_p, E_n — энергии (квадраты норм) соответственно периодической и шумовой компонент.

В то же время в работе [8] показано, что соотношение энергий периодической и шумовой компонент можно представить в виде

$$\frac{E_p}{E_n} = \frac{1 - t_0 T - r}{r}, \quad r = \frac{R(t_0)}{R(0)}. \quad (6)$$

Из (5) и (6) следует, что критерий детектирования (5) можно записать

$$\frac{R(t_0)}{R(0)} > \frac{1}{2} \left(1 - \frac{t_0}{T} \right). \quad (7)$$

Графическая интерпретация критерия (7) представлена на рис. 3. В этом случае, если пик превосходит переменный порог, принимается решение о преобладании периодической компоненты и в качестве периода основного тона выбирается значение t_0 . В противном случае принимается решение о преобладании шумовой компоненты и период основного тона не вычисляется.

Кодирование параметров

В соответствии со схемой, представленной на рис. 1, следующий этап после анализа — кодирование (квантование) параметров. Модуль квантования является, как правило, наиболее вычислительно затратным блоком вокодера. В табл. 6 и 7 представлены распределения битов на фрейм (25 мс) для вокодеров 2400 и 1200 бит/с.

Таблица 6

Параметры	Количество битов на фрейм
Линейные спектральные частоты	$4 \cdot 10 = 40$
Возбуждающий процесс	13
Коэффициент усиления	6
Синхронизационный бит	1
Итого	60

Таблица 7

Параметры	Количество битов на фрейм
Линейные спектральные частоты	16
Возбуждающий процесс	7
Коэффициент усиления	6
Синхронизационный бит	1
Итого	30

Кодирование основного тона и логарифма коэффициента усиления выполнялось с помощью равномерного квантователя на соответствующее количество битов, в то время как кодирование ЛСЧ является более сложным и существенно различается для режимов 2400 и 1200 бит/с.

Кодирование ЛСЧ для скорости 2400 бит/с. Для решения этой задачи использовался метод скалярного квантования отдельно для каждой ЛСЧ. Неравномерные уровни квантования (16 для каждой ЛСЧ) определялись с помощью алгоритма К-средних [3]. Примеры уровней квантования для первой (а) и десятой (б) ЛСЧ представлены на рис. 4.

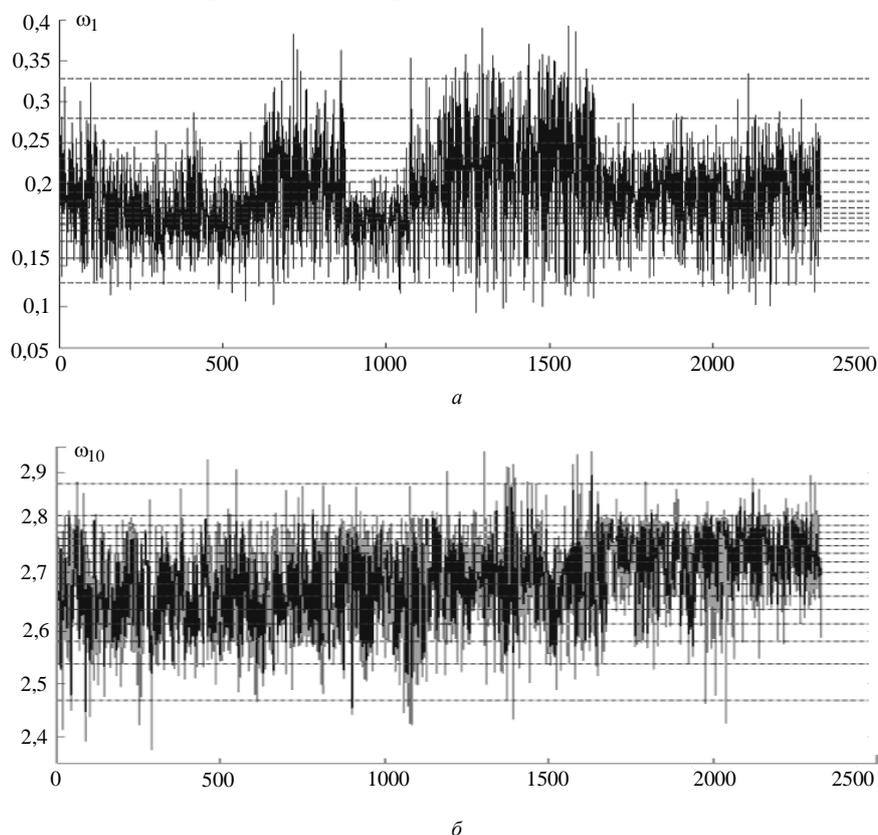


Рис. 4

Кодирование ЛСЧ для скорости 1200 бит/с. Ситуация со скоростью 1200 бит/с более сложная. Согласно табл. 7 для кодирования десяти ЛСЧ доступно всего 16 бит, что делает невозможным адекватное применение скалярного квантования ЛСЧ. Поэтому использовался принцип векторного квантования [3]. Суть состоит в кодировании множества векторов (y_1, \dots, y_p) , принадлежащих некоторому

пространству Y , множеством дискретных векторов (квантов) $(y_1^{(m)}, \dots, y_p^{(m)})$, $m = 1, \dots, M$.

В то же время прямое применение векторного квантования ЛСЧ означало бы использование базы (кодовой книги) на $2^M = 2^{16} = 65536$ 10-мерных квантов, что неприемлемо как с точки зрения ресурсов памяти ($p2^M$ элементов), так и вычислений по поиску нужного вектора. Поэтому была применена технология раздельного векторного квантования (Split Vector Quantization) [8], в которой кодируемый вектор разбивается на две подгруппы (n_1, n_2) , где $n_1 + n_2 = p=10$. При этом каждая подгруппа кодируется отдельной кодовой книгой на $2^{M/2} = 2^8 = 256$ квантов. Суммарный размер двух кодовых книг составляет $n_1 2^{M/2} + n_2 2^{M/2} = p2^{M/2}$, что существенно меньше, чем $p2^M$ (для $M = 8$ параметров отношение составляет 16 раз). Это же касается и числа вычислений по поиску оптимального вектора.

Важнейшим вопросом при построении квантователей является выбор меры близости рассматриваемых векторов. Поэтому для оценивания расстояния между двумя векторами ЛСЧ $(\omega^{(1)}$ и $\omega^{(2)})$ использовалась среднеквадратическая метрика

$$d(\omega^{(1)}, \omega^{(2)}) = \sum_{k=1}^p [\omega_k^{(1)} - \omega_k^{(2)}]^2, \quad (8)$$

а также модифицированная среднеквадратическая метрика [8]

$$d(\omega^{(1)}, \omega^{(2)}) = \sum_{k=1}^p c_k [\omega_k^{(1)} - \omega_k^{(2)}]^2, \quad c_k = [\omega_{k+1}^{(1)} - \omega_k^{(1)}]^{-1}. \quad (9)$$

Для инициализации алгоритма К-средних применялся метод неравномерной дихотомии [3]. Результаты вычисления среднеквадратического (СК) отклонения спектров [9] и меры Итакуры представлены в табл. 8 и 9 для метрик (8) и (9) соответственно.

Таблица 8

Конфигурация	Спектральное СК-искажение	Мера Итакуры
3+7	2,52	0,20
4+6	2,35	0,18
5+5	2,31	0,19
6+4	2,35	0,20
7+3	2,51	0,23

Таблица 9

Конфигурация	Спектральное СК-искажение	Мера Итакуры
3+7	2,54	0,21
4+6	2,36	0,18
5+5	2,29	0,18
6+4	2,30	0,18
7+3	2,44	0,20

соответствии с моделью речеобразования (1).

Как следует из таблиц, конфигурация (5+5) оптимальна с точки зрения погрешности квантования. Эта конфигурация также оптимальна с точки зрения вычислительных затрат по поиску ближайшего вектора (требуя вычисление $2^{M/2+1}$ норм $p/2$ -разрядных векторов).

Процедура синтеза. Процедура синтеза, как правило, проще процедуры анализа. Так, процедуры декодирования основного тона, коэффициента усиления и ЛСЧ сводятся к извлечению векторов по индексам из соответствующих кодовых книг. Имея восстановленные АР-коэффициенты, коэффициент усиления и возбуждающий процесс, на каждом фрейме речевой сигнал восстанавливается в

Практическое внедрение разработанного алгоритма

Все математические операции были закодированы на языке ассемблера процессора ADSP-2191. Для реализации функции арксинуса при подсчете ЛСЧ использовался табличный подход. Численные константы и коэффициенты фильтров хранились во внутренней памяти. Для повышения скорости выполнения операций также использовалась архитектура процессора ADSP-2191, что позволило эффективно реализовать отдельные модули вокодеров. Вышеописанные работы по оптимизации вычислений позволили реализовать вокодер в реальном времени на процессоре ADSP-2191. Время, требуемое для процедур анализа и синтеза на одном фрейме продолжительностью 25 мс, представлено в табл. 10.

Таблица 10

Вокодер	Анализ	Синтез	Анализ + синтез
Вокодер 1200 бит/с	10,1 мс	4,8 мс	14,9 мс
Вокодер 2400 бит/с	9,4 мс	5,0 мс	14,4 мс

Из табл. 10 следует, что суммарное время для выполнения процедур анализа и синтеза не превышает 60 % от продолжительности фрейма, что свидетельствует о реализуемости разработанных вокодеров в режиме реального времени

Заключение

Рассмотрена актуальная задача вычисления и кодирования параметров авторегрессионной модели речи при построении низкоскоростных кодеров речевого сигнала, работающих на скоростях 1200 и 2400 бит/с. В частности, предложен способ целочисленной реализации рекурсии Левинсона–Дарбина для расчета коэффициентов авторегрессии, который продемонстрировал снижение погрешностей их вычисления по сравнению с рекурсией Шура. Также был применен эффективный метод расчета линейных спектральных частот, основанный на свойстве межфреймовой упорядоченности и позволивший снизить средние вычислительные затраты в 3,42 раза по сравнению с широко используемым методом Кабала–Рамачандрана. Для расчета периода основного тона применялся подход, основанный на сравнении автокорреляционной функции сигнала с переменным порогом. Работы по оптимизации вычислений позволили реализовать вокодер в реальном времени на процессоре ADSP-2191.

В.Ю. Семенов

МЕТОДИ ОБЧИСЛЕННЯ ТА КОДУВАННЯ ПАРАМЕТРІВ АВТОРЕГРЕСІЙНОЇ МОДЕЛІ МОВЛЕННЯ ЗА РОЗРОБКИ ВОКОДЕРА НА БАЗІ СИГНАЛЬНОГО ПРОЦЕСОРА З ФІКСОВАНОЮ ТОЧКОЮ

Розглянуто задачу обчислення і кодування параметрів вокодерів, що працюють на швидкостях 1200 і 2400 біт/с. Запропоновано ефективні методи розрахунку і кодування параметрів авторегресійної моделі мовного сигналу, таких як лінійні спектральні частоти і період основного тону. Роботи з оптимізації обчислень дозволили реалізувати вокодер у реальному часі на сигнальному процесорі ADSP-2191.

Ключові слова: авторегресійна модель, рекурсія Левінсона–Дарбіна, лінійні спектральні частоти, період основного тону, арифметика з фіксованою комою.

V.Yu. Semenov

METHODS OF CALCULATION AND CODING
PARAMETERS OF AUTOREGRESSIONAL
SPEECH MODELS WHEN DEVELOPING
A VOCODER ON THE BASIS OF THE FIXED POINT
SIGNAL PROCESSOR

The task of calculating and coding the parameters of vocoders operating at speeds of 1200 and 2400 bps is considered. Efficient methods for calculating and coding the parameters of an autoregressive speech signal model, such as linear spectral frequencies and the period of the fundamental tone, are proposed. The computations optimization work allowed real-time vocoder to be implemented on the ADSP-2191 signal processor.

Keywords: autoregressive model, Levinson-Durbin recursion, line spectral frequencies, pitch period, fixed-point arithmetic.

1. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов. М. : Радио и связь. 1981. 420 с.
2. McCree A.V., Barnwell T.P. A mixed excitation LPC vocoder model for low bit rate speech coding. *IEEE Trans. Speech Audio Processing*. 1995. **42**. P. 242–249. DOI : 10.1109/89.397089.
3. Макхоул Дж., Русос С., Гиш Г. Векторное квантование при кодировании речи. *ТИИЭР*. 1985. **73**. С. 19–61. DOI : 10.1109/PROC.1985.13340.
4. Kabal P., Ramachandran R. P. The computation of line spectral frequencies using Chebyshev polynomials. *IEEE Trans. Acoust. Speech Signal Processing*. 1980. **28**. P. 562–574. DOI : 10.1109/TASSP.1986.1164983.
5. Семенов В.Ю. Метод нахождения всех корней системы нелинейных алгебраических уравнений, основанный на операторе Кравчика. *Кибернетика и системный анализ*. 2015. № 5. С. 169–175. DOI : 10.1007/s10559-015-9775-0.
6. Semenov V. Computation of immittance and line spectral frequencies based on interframe ordering property. *Journal of Computers*. 2007. **2**. N 7. P. 75–80
7. Вовк И.В., Семенов В.Ю. Автоматическое обнаружение и распознавание сухих хрипов на основе анализа их автокорреляционной функции. *Акустичний вісник*. 2005. **8**. №3. С. 17–23.
8. Paliwal K.K., Atal B.S. Efficient vector quantization of LPC parameters at 24 bits/frame. *IEEE Trans. Speech Audio Proces.* 1993. **1**. N 1. P. 3–14. DOI : 10.1109/ICASSP.1991.150426.

Получено 11.05.2018

После доработки 28.08.2018