ИНФОРМАТИКА

2009 **апрель-июнь** **№ 2**

ОБРАБОТКА СИГНАЛОВ И РЕЧИ

УДК 621.391.7

А.А. Борискевич, А.О. Рак

ВЕКТОРНО-ПАРАМЕТРИЧЕСКОЕ НИЗКОСКОРОСТНОЕ СЖАТИЕ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ СУПЕРКАДРОВ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Разрабатывается алгоритм векторно-параметрического низкоскоростного сжатия речи, основанный на использовании параметрической модели синтеза речевого сигнала с линейным предсказанием, суперкадров с переменной структурой, векторного квантования параметров суперкадра (коэффицента усиления, периода основного тона и LSF(line spectrum frequency)-коэффициентов) и интерполяции LSF-кадров. Даются рекомендации по выбору структуры суперкадра в зависимости от типа передаваемых параметров модели речевого сигнала. Осуществляется программная реализация алгоритма низкоскоростного параметрического сжатия речи в среде моделирования Matlab. Показывается, что разборчивость речи сохраняется при битовых скоростях 300–800 бит/с. Устанавливается, что увеличение битовой скорости обычно не приводит к значительному улучшению качества звучания из-за ограничений, накладываемых выбранной моделью речеобразования.

Введение

В настоящее время в технике сжатия речевого сигнала (РС) используются в основном параметрически-адаптивные алгоритмы кодирования речи на основе различных модификаций метода линейного предсказания, которые заключаются в различном представлении сигнала возбуждения [1–9]. Анализ современных низкоскоростных кодеков РС показывает, что актуальным является качественное и компактное представление параметров, характеризующих передаточную характеристику голосового тракта и сигнала возбуждения.

Главными качественными характеристиками низкоскоростных параметрических вокодеров являются скорость выходного битового потока и разборчивость синтезированной речи. В случае необходимости обеспечения высокой крипто- и помехозащищенности, а также в системах связи с морскими судами или по узкополосным КВ-радиоканалам в условиях помех и замираний требуются вокодеры со скоростью битового потока ниже 600–800 бит/с, вплоть до 300 бит/с. Анализ современных кодеков [2–9] указывает на однозначную связь качества синтезируемой речи на низких скоростях кодирования со степенью их адаптации и перспективность перехода систем кодирования речи к многопараметрической адаптации в условиях априорной неопределенности в описании моделей и внешней среды функционирования кодеков. По этой причине для сверхнизкоскоростной компрессии речи были разработаны другие подходы: вокодеры с обработкой нескольких кадров фиксированной длины, объединенных в суперкадр, вокодеры с переменной длиной кадра на основе сегментации речи и фонемные вокодеры [4–7]. Первые используют векторное квантование траекторий параметров для всего суперкадра и обеспечивают повышение качества кодирования за счет динамического перераспределения информационных бит между квантуемыми параметрами и кадрами, входящими в суперкадр. Вторые сегментируют речь и описывают большие однородные речевые фрагменты целиком в пределах естественных границ, а не отдельные небольшие кадры с фиксированными границами. Третьи используют методы теории распознавания образов для выделения элементов звукового алфавита из текущей речи на передающей стороне с последующим синтезом по алфавиту на приемной стороне. Во всех случаях становится возможным создавать приемлемое по точности описание для фрагментов речи длительностью около 100 мс с использованием 40–80 бит на фрагмент.

Наибольшее развитие в мировой практике получили вокодеры первого типа, использующие суперкадр. В настоящее время за рубежом существуют военные стандарты MIL-STD-3005

и NATO STANAG 4591 на низкоскоростной вокодер MELPe (enhance mixed excitation linear predictive) со скоростями 600, 1200 и 2400 бит/с [3], а на вокодеры со скоростью битового потока 300 бит/с объявлен международный конкурс [8]. Однако такие стандарты в силу специфики применения являются закрытыми. Поэтому разработка собственных алгоритмов компрессии речи со скоростями 300–800 бит/с представляет большой научный и практический интерес. С коммерческой точки зрения разработка оригинальных алгоритмов также желательна для обеспечения независимости от держателей патентов стандартных решений.

Целью работы является разработка алгоритма сверхнизкоскоростного сжатия, обеспечивающего диапазон скоростей битового потока 300–800 бит/с при сохранении минимально допустимой разборчивости синтезированной речи за счет повышения степени его адаптации.

1. Описание алгоритма векторно-параметрического низкоскоростного сжатия РС

Особенностью разработанного алгоритма, основанного на раздельном описании голосового тракта и сигнала возбуждения, является временная структура суперкадра, кадры в котором имеют различную длину в зависимости от определяемого параметра модели речеобразования: LSF-параметра, моделирующего состояния голосового тракта, а также Pitch-параметра и Gain-параметра, моделирующих состояния голосовой щели.

Выбор данной структуры основан на предположении, что для определения речевых параметров требуется различная длина кадра и между соседними LSF-кадрами суперкадра существует достаточно сильная корреляция, обусловленная малыми изменениями в конфигурации голосового тракта человека.

Известно, что информационная ценность звукового сигнала на длительности отдельного звукового объекта, например фонемы, различна [10]. Развитие звукового объекта во времени включает три фазы: атаки (крутого нарастания или переднего фронта), поддержки и мягкого спада (заднего фронта). Особую ценность представляют участки начала звучания звукового объекта (атаки), определяющие основную смысловую информативность РС. Устранение атак из РС приводит к его полной неразборчивости. В то же время сохранение только атак позволяет обеспечить словесную разборчивость на уровне 85 %. В РС атаки составляют около 15 % длительности сигнала. Длительность атаки для РС составляет от 2 до 40 мс (согласные звуки) [10]. В связи с этим для повышения разборчивости синтезированной речи минимальная длительность Gain-кадра может составлять 2 мс и более, что позволит более точно отразить кратковременную динамику изменения звуковых объектов в РС. Кроме того, данный способ разбиения позволяет лучше передать низкоуровневые согласные звуки, в основном определяющие разборчивость сообщения.

Для повышения точности определения частоты основного тона размер Pitch-кадров должен быть достаточно большим, чтобы более точно отразить долговременную динамику (периодический характер звуковых объектов), учитывая диапазон изменения Pitch f_P у разных людей (мужчин, женщин, детей) от $f_{P\min} = 50$ до $f_{P\max} = 400$ Гц [11]. В связи с этим длительность Pitch-кадров выбирается из соотношения $T_{N_L} = \frac{1000 N_L}{f_S} \geq \frac{1000}{f_{P min}}$, где N_L – число отсчетов в Pitch-кадре, f_S – частота дискретизации, и составляет порядка 20 мс и выше.

Выбор размера LSF-кадра обусловлен соблюдением условий линейности и постоянства параметров голосового тракта, которые в основном определяются его инерционностью. Известно, что голосовой тракт человека можно рассматривать как линейную систему с постоянными параметрами в промежутке времени 10–30 мс [9]. В связи с этим длительность LSF-кадра может изменяться от 10 до 30 мс.

Из-за межкадровой корреляции LSF-параметры могут передаваться не для каждого кадра. Непереданные значения на приемной стороне интерполируются. Данный прием позволяет не только значительно снизить скорость битового потока, но и уменьшить объем вычислений, так как непередаваемые значения можно не вычислять [9]. Использование кадров различной длины для определения речевых параметров позволяет одновременно с понижением скорости битового потока добиться улучшения качества синтезируемой речи.

Алгоритм векторно-параметрического низкоскоростного сжатия РС на основе суперкадров с переменной структурой состоит из следующих шагов:

Шаг 1. Формирование m-го суперкадра РС.

Из условия достижения требуемой битовой скорости $R = B / T_{SF}$ бит/с исходный РС $\hat{x}(n)$ разбивается на неперекрывающиеся суперкадры $\hat{x}_m(n) = x(n + mN_M)$. Здесь B – заданное число бит на суперкадр; $T_{SF} = N_M / f_S$ – длительность суперкадра, с; f_S – частота дискретизации; $n = \overline{0, N_M - 1}$; N_M – число отсчетов m-го суперкадра; $m = \overline{0, M - 1}$; m – номер суперкадра; M – число суперкадров.

Шаг 2. Предыскажение m-го кадра РС.

Для улучшения качества синтезированной речи может применяться предыскажающий фильтр. РС взвешивается корректирующим фильтром с амплитудно-частотной характеристикой, имеющей подъем в высокочастотной области, для улучшения передачи более слабых высокочастотных компонент РС, определяющих в основном его разборчивость [9].

Форма представления фильтра коррекции первого порядка во временной области имеет вид

$$x_m(n) = \hat{x}_m(n) - k_d \hat{x}_m(n-1), \quad (1)$$

где $x_m(n)$, $\hat{x}_m(n)$ и $\hat{x}_m(n-1)$ – текущая искаженная выходная, входная и предыдущая входная выборки РС; k_d – коэффициент, характеризующий степень искажения значений выборок РС и принимающий значения в диапазоне $[0,9 - 1,0]$.

В практических приложениях обычно используется значение $k_d = 0,95$ при $f_S = 8$ кГц, а при представлении чисел в форме с фиксированной запятой часто применяется значение $k_d = 15/16 = 0,9375$.

Шаг 3. Распараллеливание и сегментация m-го суперкадра.

Для улучшения качества синтезируемой речи в соответствии с принципом параметрической адаптации суперкадры разбиваются на кадры различной длины $\hat{x}_{ml}(n) = x(n + mN_M + lN_L)$, где $l = \overline{1, L}$ – номер кадра в суперкадре; $N_L = N_M / L$ – число отсчетов в l-м кадре m-го суперкадра; L – число кадров в суперкадре.

Для достижения требуемой битовой скорости возможно отбрасывание LSF-кадров, значения которых на приемной стороне будут восстанавливаться с помощью линейной интерполяции.

Шаг 4. Вычисление LP-вектора l-го кадра m-го суперкадра.

Для вычисления вектора $\boldsymbol{\alpha}_{ml}^P$ LP(linear prediction)-коэффициентов l-го кадра m-го суперкадра, эффективно аппроксимирующих спектральную огибающую РС, используется рекурсивный алгоритм Левинсона – Дарбина [12]: $\boldsymbol{\alpha}_{ml}^P = (\alpha_{1ml}, ..., \alpha_{jml}, ..., \alpha_{Pml})^T$, где α_{jml} – j-й LP-коэффициент предсказания l-го кадра m-го суперкадра; P – порядок кратковременного предсказания.

В полосе частот до 4 кГц может быть до 4–5 формант. Следовательно, для качественного восстановления любого речевого сегмента требуется 8–10 LP-коэффициентов, позволяющих более точно аппроксимировать формантные области Фурье-спектра кадра РС [12].

Шаг 5. Преобразование LP-вектора в LSF-вектор l-го кадра m-го суперкадра.

LSF-представление РС является одним из эффективных описаний спектральной огибающей РС или состояния речевого тракта из-за следующих свойств LSF-коэффициентов [13]:

1. Свойство квантования: малые ошибки квантования в LSF-коэффициентах вызывают малые спектральные ошибки, что гарантирует стабильность синтезирующего фильтра на приемной стороне вокодера.

С учетом психоакустического восприятия человека низкочастотные LSF-коэффициенты квантуются более точно, чем более высокочастотные LSF-коэффициенты, из-за высокой чувствительности слухового анализатора человека к малым изменениям в области более низких частот.

2. Корреляционные свойства: имеется внутрикадровая и межкадровая корреляция LSF-коэффициентов.

Свойство упорядочивания указывает на то, что LSF-коэффициенты коррелируют в пределах кадра. Сильная корреляция между LSF-коэффициентами соседних кадров обусловлена малыми изменениями в конфигурации голосового тракта человека.

3. Спектральные свойства: LSF-коэффициенты характеризуют тесную связь с формантными частотами или с перцептивно важными пиками спектральной огибающей РС.

Свойства группы LSF-коэффициентов характеризуют формантную частоту и ширину формантной полосы, зависящую от близости LSF-коэффициентов в группе.

LSF-коэффициенты определяются как угловая позиция ω комплексных корней z, лежащих на единичной окружности в z-плоскости, двух LSF-полиномов $F_s(z)$ и $F_a(z)$:

$$\omega = \arg z, \tag{2}$$

где $F_s(z) = A_P(z) + z^{-P-1}A_P(z^{-1})$ и $F_a(z) = A_P(z) - z^{-P-1}A_P(z^{-1})$ – симметричный и асимметричный LSF-полиномы; $A_P(z) = 1 - \sum_{j=1}^{P} \alpha_j z^{-1}$ – LP-полином, связанный с передаточной характеристикой $H_P(z) = 1/A_P(z)$ синтезирующего фильтра; $z = |z|\exp(j\omega)$ – комплексные корни (нули) LSF-полинома; $|z| = 1$ – радиус-вектор единичной окружности.

Порядок расположения LSF-коэффициентов в частотной области

$$0 < \omega_{1F_s} < \omega_{1F_a} < \ldots < \omega_{(P/2)F_s} < \omega_{(P/2)F_a} < \pi, \tag{3}$$

где ω_{jF_s} и ω_{jF_a} – линейные спектральные пары, которые соответствуют двум соседним корням, принадлежащим двум LSF-полиномам: нечетные угловые частоты соответствуют корням $F_s(z)$, а четные угловые частоты – корням $F_a(z)$.

Конечным результатом данного шага является P-мерный LSF-вектор $\boldsymbol{\omega}_{ml} = (\omega_{1ml}, \ldots, \omega_{Pml})^T$ l-го кадра m-го суперкадра.

Шаг 6. Вычисление периода основного тона l-го кадра m-го суперкадра.

Для определения периода основного тона используется следующее соотношение:

$$T_{Pml} = \arg\max_n C'_{ml}(n) \tag{4}$$

при $\max_n C'_{ml}(n) \geq T_{nop}$.

Здесь T_{nop} – пороговое значение кепстра, при превышении которого l-й кадр m-го суперкадра классифицируется как вокализованный, а координата пика дает достаточно хорошую оценку периоду основного тона [14]. Если максимум кепстра не превышает порога, то l-й кадр m-го суперкадра классифицируется как невокализованный;

$C'_{ml}(n) = \frac{1}{N_L} l_P(n) \sum_{k=0}^{N_L-1} X'_{ml}(k)\exp(j\omega_k n)$ – значения вещественного кепстра l-го кадра m-го суперкадра, характеризующегося наличием импульсов (резких пиков) в точках, кратных периоду основного тона для вокализованного кадра, или отсутствием ярко выраженных пиков для невокализованного кадра при больших значениях временных индексов n, где $l_P(n) = \begin{cases} 0 & \text{при } n < f_s / f_{P\max}, \\ 1 & \text{при } f_s / f_{P\max} \leq n \leq f_s / f_{P\min} \end{cases}$ – оконная функция для выделения только тех значений вещественного кепстра, которые характеризуют частоту основного тона; $X'_{ml}(k) = \log|X_{ml}(k)|$ –

логарифм модуля Фурье-спектра l-го кадра m-го суперкадра; $X_{ml}(k)=\sum_{n=0}^{N_M-1} x_M w(n)\exp(-j\omega_k n)$ – комплексное значение k-й спектральной составляющей Фурье-спектра l-го кадра m-го суперкадра; $w(n)$ – оконная весовая функция Хэмминга [15], минимизирующая влияние краевых эффектов на точность вычисления значений спектральных составляющих; $\omega_k = 2\pi k/N_L$ – частота k-й спектральной составляющей Фурье-спектра l-го кадра m-го суперкадра; $|X_{ml}(k)| = \left(X_{ml}(k)X^*_{ml}(k)\right)^{1/2}$ – значение k-й спектральной составляющей амплитудного Фурье-спектра l-го кадра m-го суперкадра.

Шаг 7. Вычисление коэффициента усиления l-го кадра m-го суперкадра.

Коэффициент усиления для вокализованных и невокализованных кадров различен. Признак вокализованности кадра определяется по значению Pitch анализируемого кадра: ненулевым значениям Pitch соответствуют вокализованные кадры.

Для вокализованного и невокализованнго l-го кадра m-го суперкадра коэффициент усиления g_{ml} соответственно определяется с помощью соотношений

$$g_{ml} = \frac{T_{Pml}}{N_L}\sum_{n=0}^{N_L-1} x^2_{ml}(n); \tag{5}$$

$$g_{ml} = \frac{1}{N_L}\sum_{n=0}^{N_L-1} x^2_{ml}(n), \tag{6}$$

где T_{Pml} – период основного тона l-го кадра m-го суперкадра, используемый в качестве коэффициента для компенсации ослабления пиков, вызванного усреднением.

Шаг 8. Кодирование LSF-вектора l-го кадра m-го суперкадра.

Обучающая последовательность, состоящая из большого количества векторов, была выбрана из коллекции речевых выборок, моделирующих различные голосовые и шумовые характеристики для обеспечения лучшего качества реконструированной речи не только по объективным, но и субъективным оценкам. Выбран итерационный LBG-алгоритм [16] для кластеризации множества обучающих векторов в множество векторов кодовой книги.

Для достижения компромисса между битовыми затратами, искажениями квантования, вычислительной сложностью квантования и требуемым объемом памяти для хранения кодовых книг используется векторный квантователь с расщеплением (ВКР) и многоуровневый векторный квантователь (МУВК) [9, 16–19]. Данные квантователи позволяют достичь спектральной прозрачности квантования при более низких скоростях кодирования.

При использовании ВКР результирующая I-элементная и P-мерная кодовая книга $C=\left\{\tilde{\boldsymbol{\omega}}_i \middle| \tilde{\boldsymbol{\omega}}_i \in R^P, i=1,2,...,I\right\}$ состоит из K независимых кодовых книг:

$$C=\{C_k : k\in K\}, \tag{7}$$

где $C_k=\left\{\tilde{\boldsymbol{\omega}}_{ki}, \tilde{\boldsymbol{\omega}}_{ki}\in R_k^{P_k}, i=1,2,...,I_k\right\}$, $\tilde{\boldsymbol{\omega}}_{ki}=\left(\tilde{\omega}_{ki1},...,\tilde{\omega}_{kij},...,\tilde{\omega}_{kiP_k}\right)^T$, $I_k=2^{b_k}$ и K – i-й P_k-мерный кодовый вектор LSF-коэффициентов кодовой книги C_k, размер b_k-битовой кодовой книги C_k и число расщеплений соответственно; $k=\overline{1,K}$; $i=\overline{1,I_k}$; $j=\overline{1,P_k}$; $I=\sum_{k=1}^{K} I_k$, $P=\sum_{k=1}^{K} P_k$.

Из (7) видно, что исходное P-мерное векторное пространство R^P делится на K субпространств меньшей размерности $\left\{R_k^{P_k}\right\}_{k=1}^{K}$, т. е. i-й кодовый LSF-вектор $\tilde{\boldsymbol{\omega}}_i = \left(\tilde{\omega}_{i1},...,\tilde{\omega}_{ij},...,\tilde{\omega}_{iP}\right)^T \in R^P$ расщепляется, или делится, на K LSF-субвекторов размерности P_k. Каждая из кодовых книг C_k, использующих b_k бит/субвектор при условии, что $b = \sum_{k=1}^{K} b_k$, где b – число бит на LSF-вектор кадра, строятся независимо друг от друга. Данные кодовые книги обучаются посредством использования соответствующих субвекторов выбранного обучающего множества. Заметим, что выбор P_k и b_k обусловлен свойствами LSF-коэффициентов. В данном алгоритме LSF-вектор $\tilde{\omega}_i$ делится на три субвектора с учетом их психоакустической значимости: первый субвектор $\tilde{\boldsymbol{\omega}}_{1i} = \left(\tilde{\omega}_{1i1},\tilde{\omega}_{1i2},\tilde{\omega}_{1i3}\right)^T$ при $P_1 = 3$ содержит три самых низкочастотных LSF-коэффициента, второй $\tilde{\boldsymbol{\omega}}_{2i} = \left(\tilde{\omega}_{2i1},\tilde{\omega}_{2i2},\tilde{\omega}_{2i3}\right)^T$ при $P_2 = 3$ – три среднечастотных LSF-коэффициента и третий $\tilde{\boldsymbol{\omega}}_{3i} = \left(\tilde{\omega}_{3i1},\tilde{\omega}_{3i2},\tilde{\omega}_{3i3},\tilde{\omega}_{3i4}\right)^T$ при $P_3 = 4$ – четыре самых высокочастотных LSF-коэффициента.

В этом случае результирующий кодовый LSF-вектор $\tilde{\boldsymbol{\omega}}_{ml}$ исходного LSF-вектора $\boldsymbol{\omega}_{ml}$ l-го кадра m-го суперкадра является конкатенацией трех субвекторов и определяется соотношением

$$\tilde{\boldsymbol{\omega}}_{ml} = \left\{\tilde{\boldsymbol{\omega}}_{1ml},\tilde{\boldsymbol{\omega}}_{2ml},\tilde{\boldsymbol{\omega}}_{3ml}\right\}. \tag{8}$$

В качестве меры искажений для определения оптимального множества кодовых векторов и выбора оптимального кодового вектора из кодовой книги C_k с целью передачи входного вектора используется квадрат евклидова LSF-расстояния:

$$d_{k\,LSF}(\boldsymbol{\omega}_{kl},\tilde{\boldsymbol{\omega}}_{kl}) = \sum_{j=1}^{P_k}\left[(\omega_{kl\,j} - \tilde{\omega}_{kl\,j})\right]^2. \tag{9}$$

Из-за ограниченной структуры кодовой книги характеристики ВКР являются субоптимальными, так как векторные квантователи строятся независимо в пределах каждого субпространства меньшей размерности.

МУВК является субоптимальным из-за ограниченной структуры кодовой книги, а также последовательной природы построения и поиска кодовых векторов. Однако по сравнению с ВКР он имеет большую гибкость в понятиях сложности поиска, хранения кодовой книги и защиты от ошибок. Это обусловлено тремя его возможностями:

– учета долговременной линейной и нелинейной корреляции между компонентами векторов;

– использования более высокой размерности кодовых векторов для уменьшения искажений квантования при сохранении битового ресурса или уменьшения битового ресурса при сохранении искажений посредством выбора оптимальной геометрической формы ячеек (областей кодирования) кодовой книги;

– выбора различных размеров ячеек для повышения точности аппроксимации функции плотности вероятности распределения значений РС.

Кодовая книга МУВК состоит из индивидуальных кодовых книг с различным числом кодовых векторов для управления соотношением битовая скорость/искажения на каждой стадии квантования.

Использование МУВК обеспечивает последовательное аппроксимационное квантование исходных LSF-векторов. Результирующий кодовый LSF-вектор $\tilde{\boldsymbol{\omega}}_{ml}$ исходного LSF-вектора $\boldsymbol{\omega}_{ml}$ l-го кадра m-го суперкадра при S-уровневом квантовании определяется соотношением

$$\tilde{\boldsymbol{\omega}}_{ml} = \psi\left(\tilde{\boldsymbol{\omega}}_{1ml},...,\Delta\tilde{\boldsymbol{\omega}}_{kml},...,\Delta\tilde{\boldsymbol{\omega}}_{Sml}\right) = \tilde{\boldsymbol{\omega}}_{1ml} + \sum_{k=1}^{S-1}\Delta\tilde{\boldsymbol{\omega}}_{kml}\ , \tag{10}$$

где $\tilde{\boldsymbol{\omega}}_{1ml}$ – квантованный P-мерный кодовый вектор первого уровня I-элементной и P-мерной кодовой книги C_1; $\Delta\tilde{\boldsymbol{\omega}}_{kml}$ – квантованный P-мерный вектор ошибки на k-м уровне, показывающий, на сколько исходный P-мерный вектор $\boldsymbol{\omega}_{ml}$ отличается от P-мерного вектора $(k-1)$-го уровня $\tilde{\boldsymbol{\omega}}_{(k-1)ml} = \tilde{\boldsymbol{\omega}}_{1ml} + \sum_{\lambda=1}^{k-1}\Delta\tilde{\boldsymbol{\omega}}_{\lambda ml}$, взятого из кодовой книги C_{k-1}. На данном шаге используется трехуровневый ($S=3$) векторный квантователь.

Конечным результатом данного шага является кодирование LSF-вектора l-го кадра m-го суперкадра одним из множеств индексов: $\tilde{\boldsymbol{\omega}}_{ml} \Rightarrow \left\{i_{1m\omega l}, i_{2m\omega l}, i_{3m\omega l}\right\}$ или $\tilde{\boldsymbol{\omega}}_{ml} \Rightarrow \left\{i_{1m\,\omega l}, i_{2m\,\Delta\omega l}, i_{3m\Delta\omega l}\right\}$.

Шаг 9. Кодирование вектора периодов основного тона m-го суперкадра.

Использование суперкадров позволяет объединять значения периодов основного тона кадров суперкадра в вектор $\mathbf{T}_m = (T_{m1},...,T_{ml.},...,T_{mL})$ и применять к нему векторное квантование для их компактного описания. В качестве меры искажений для определения оптимального множества кодовых векторов и выбора оптимального кодового вектора из кодовой книги используется квадрат евклидова расстояния

$$d_{mT} = \sum_{l=1}^{L}\left(T_{ml} - \tilde{T}_{ml}\right)^2 . \tag{11}$$

Конечным результатом данного шага является кодирование вектора периодов основного тона m-го суперкадра индексом i_{mT}: $\mathbf{T}_m \Rightarrow i_{mT}$.

Шаг 10. Кодирование вектора коэффициентов усиления m-го суперкадра.

Значения коэффициентов усиления суперкадра объединяются в вектор $\mathbf{g}_m = (g_{m1},..., g_{ml},..., g_{mL})$ с целью его компактного описания. В качестве меры искажений для определения оптимального множества кодовых векторов и выбора оптимального кодового вектора из кодовой книги используется квадрат евклидова расстояния

$$d_{mg} = \sum_{l=1}^{L}\left(g_{ml} - \tilde{g}_{ml}\right)^2 . \tag{12}$$

Конечным результатом данного шага является кодирование вектора коэффициентов усиления m-го суперкадра индексом i_{mg}: $\mathbf{g}_m \Rightarrow i_{mg}$.

Шаг 11. Формирование битового потока m-го суперкадра.

Суть векторного квантования заключается в кодировании векторов параметров РС индексами кодовых книг и последующем восстановлении векторов параметров РС по переданным индексам на приемной стороне. В связи с этим по каналу связи передается множество индексов каждого суперкадра РС: $\left\{\left\{i_{1m\omega l}, i_{2m\omega l}, i_{3m\omega l}\right\}_{l=1}^{L}, i_{mT}, i_{mg}\right\}_{m=1}^{M}$ для ВКР и $\left\{\left\{i_{1m\,\omega l}, i_{2m\,\Delta\omega l}, i_{3m\Delta\omega l}\right\}_{l=1}^{L}, i_{mT}, i_{mg}\right\}_{m=1}^{M}$ для МУВК.

2. Описание алгоритма векторно-параметрического синтеза m -го кадра РС на основе суперкадров с переменной структурой

Векторное кодирование и декодирование являются асимметричными: кодер осуществляет поиск кодовых векторов, а декодер – просто табличный поиск. Поэтому алгоритм синтеза речи требует меньше временных ресурсов на декодирование параметров РС, чем алгоритм векторного кодирования параметров РС на передающей стороне, и состоит из следующих шагов:

Шаг 1. Распределение битового потока для восстановления параметров РС.

Шаг 2. Восстановление квантованных значений LSF-вектора l -го кадра m -го суперкадра.

Кодовый LSF-вектор $\tilde{\boldsymbol{\omega}}_{ml}$ исходного LSF-вектора $\boldsymbol{\omega}_{ml}$ l-го кадра m -го суперкадра восстанавливается с помощью переданного множество индексов ВКР:

$$\left\{i_{1m\omega l}, i_{2m\omega l}, i_{3m\omega l}\right\} \Rightarrow \tilde{\boldsymbol{\omega}}_{ml}. \tag{13}$$

Шаг 3. Интерполирование LSF-векторов m -го суперкадра.

Учет межкадровой корреляции между LSF-кадрами суперкадра позволяет передавать значения LSF-коэффициентов лишь для определенных кадров. Непереданные LSF-кадры восстанавливаются на приемной стороне с помощью линейной интерполяции, что позволяет значительно уменьшить объем передаваемого битового потока:

$$\tilde{\boldsymbol{\omega}}_{ml} = (1-\alpha_l)\tilde{\boldsymbol{\omega}}_{m0} + \alpha_l \tilde{\boldsymbol{\omega}}_{m(L-1)} \tag{14}$$

при $0 \le l \le L-1$, где $\tilde{\boldsymbol{\omega}}_{ml}$ – интерполяционный LSF-вектор для l -го кадра m -го суперкадра; L – интерполяционный коэффициент или число кадров в m -м суперкадре; $\alpha_l = l/(L-1)$; $\tilde{\boldsymbol{\omega}}_{m0}$ и $\tilde{\boldsymbol{\omega}}_{m(L-1)}$ – переданные LSF-векторы первого и последнего кадра m -го суперкадра соответственно.

В случае передачи одного LSF-вектора остальные LSF-векторы текущего суперкадра интерполируются с использованием соответствующего LSF-вектора из предыдущего суперкадра. Конечным результатом данного шага является множество P -мерных LSF-векторов $\left\{\tilde{\boldsymbol{\omega}}_{ml} = (\tilde{\omega}_{1ml}, ..., \tilde{\omega}_{Pml})^T\right\}_{l=1}^{L}$ m -го суперкадра.

Шаг 4. Преобразование LSF-коэффициентов в LP-коэффициенты m -го суперкадра.

Преобразование из LP-коэффициентов в LSF-коэффициенты является обратимым, т. е. можно точно вычислить LP-коэффициенты из LSF-коэффициентов [13]. Перевод LSF-коэффициентов в LP-коэффициенты осуществляется с использованием соотношения

$$A_P(z) = \frac{F_S(z) + F_A(z)}{2}. \tag{15}$$

Конечным результатом данного шага является множество P -мерных LP-векторов $\left\{\tilde{\boldsymbol{\alpha}}_{ml}^{P} = (\tilde{\alpha}_{1ml}, ..., \tilde{\alpha}_{jml}, ..., \tilde{\alpha}_{Pml})^T\right\}_{l=1}^{L}$ m -го суперкадра.

Шаг 5. Восстановление квантованных значений вектора периодов основного тона m -го суперкадра.

Кодовый вектор $\tilde{\boldsymbol{T}}_m$ исходного вектора $\boldsymbol{T}_m$ m -го суперкадра восстанавливается с помощью переданного индекса i_{mT} кодовой книги векторов $\left\{\tilde{T}_m\right\}$:

$$i_{mT} \Rightarrow \tilde{\boldsymbol{T}}_m = (\tilde{T}_{m1}, ..., \tilde{T}_{ml}, ..., \tilde{T}_{mL}). \tag{16}$$

Шаг 6. Формирование сигнала возбуждения для синтеза вокализованного $e_{Tml}(n)$ или невокализованного e_{ml} l -го кадра m -го суперкадра:

$$e_{Tml}(n) = e_T(n + mN_M + lN_L); \tag{17}$$

$$e_{ml}(n) = e(n + mN_M + lN_L). \tag{18}$$

Шаг 7. Восстановление квантованных значений вектора коэффициентов усиления m-го суперкадра.

Кодовый вектор $\widetilde{\boldsymbol{g}}_m$ исходного вектора $\boldsymbol{g}_m$ m-го суперкадра восстанавливается с помощью переданного индекса i_{mg} кодовой книги векторов $\{\tilde{\boldsymbol{g}}_m\}$:

$$i_{mg} \Rightarrow \tilde{\boldsymbol{g}}_m = (\tilde{g}_{m1},...,\tilde{g}_{ml.},...,\tilde{g}_{mL}). \tag{19}$$

Шаг 8. Синтез вокализованного $x'_{Tm}(n)$ или невокализованного $x'_m(n)$ m-го суперкадра.

Стабильность синтезирующего фильтра, которая является важным предварительным требованием для кодирования речи, гарантируется посредством квантования LP-коэффициентов в LSF-области:

$$x'_{Tm}(n) = \sum_{l_g=1}^{L_g} \tilde{g}_{ml_g} e_{Tml_g}(n) * h_{Tml_g}(n); \tag{20}$$

$$x'_m(n) = \sum_{l_g=1}^{L_g} \tilde{g}_{ml_g} e_{ml_g}(n) * h_{ml_g}(n), \tag{21}$$

где $*$ – символ обозначения линейной дискретной свертки; l_g и L_g – текущий номер Gain-кадра и число Gain-кадров в m-м суперкадре; $h_{Tml_g}(n)$ и h_{ml_g} – импульсные характеристики синтезирующего всеполюсного фильтра с ограниченным порядком P для вокализованного и невокализованного l_g-го кадра m-го суперкадра соответственно.

Шаг 9. Коррекция предыскажений вокализованного $x'_{Tm}(n)$ или невокализованного $x'_m(n)$ m-го суперкадра.

Для компенсации предыскажений, введенных на шаге 2 алгоритма векторно-параметрического низкоскоростного сжатия РС, используются следующие соотношения:

$$\hat{x}'_{Tm}(n) = x'_{mT}(n) + k_d x'_{Tm}(n-1); \tag{22}$$

$$\hat{x}'_m(n) = \hat{x}'_m(n) + k_d x'_m(n-1). \tag{23}$$

3. Результаты моделирования

Проведены эксперименты для различных РС, кодируемых со скоростями от 800 до 300 бит/с. Поскольку низкоскоростные вокодеры не сохраняют форму сигнала, использование в качестве критерия качества восстановления речи во временной области среднеквадратической ошибки невозможно.

Для оценки влияния компактного представления параметров, характеризующих передаточную характеристику голосового тракта, сигнала возбуждения и переменную структуру суперкадра, на качество синтезированного сигнала используются пять мер искажений речи [20], основанных на применении амплитудной информации, так как слуховая система человека является относительно нечувствительной к фазовой информации.

Объективная мера искажения Itakura – Saito (IS) m-го кадра синтезированного сигнала $y(n)$ задается соотношением [20]

$$d_{IS}(m) = \left(\frac{\sigma_x^2(m)}{\sigma_y^2(m)}\right)\left(\frac{\boldsymbol{\alpha}_y(m) R_x(m) \boldsymbol{\alpha}_y^T(m)}{\boldsymbol{\alpha}_x(m) R_x(m) \boldsymbol{\alpha}_x^T(m)}\right) + \log\left(\frac{\sigma_y^2(m)}{\sigma_x^2(m)}\right) - 1, \tag{24}$$

где $\boldsymbol{\alpha}_x(m)=(1,\alpha_{x1},\ldots,\alpha_{xP})^T$ и $\boldsymbol{\alpha}_y(m)=(1,\alpha_{y1},\ldots,\alpha_{yP})^T$ – векторы LPC-коэффициентов m-го кадра обработанных исходного и синтезированного сигналов; $R_x(m)$ – автокорреляционная матрица m-го кадра обработанного исходного сигнала; $\sigma_x^2(m)=R_x(m)\boldsymbol{\alpha}_x^T(m)$ и $\sigma_y^2(m)=R_y(m)\boldsymbol{\alpha}_y^T(m)$ – коэффициенты усиления m-го кадра обработанных исходного и синтезированного сигналов соответственно; $m=\overline{1,M}$; M – число кадров в сигнале.

Мера логарифмического правдоподобия [20]

$$d_{LLR}(m)=\log\left(\frac{\boldsymbol{\alpha}_y(m)R_x(m)\boldsymbol{\alpha}_y^T(m)}{\boldsymbol{\alpha}_x(m)R_x(m)\boldsymbol{\alpha}_x^T(m)}\right). \tag{25}$$

Мера логарифмического отношения площадей [20]

$$d_{LAR}(m)=\left|\frac{1}{P}\sum_{p=1}^{P}\left(\log\frac{1+r_{xp}(m)}{1-r_{xp}(m)}-\log\frac{1+r_{yxp}(m)}{1-r_{yp}(m)}\right)^2\right|^{0,5}, \tag{26}$$

где $r_{xp}(m)$ и $r_{yp}(m)$ – p-е коэффициенты отражения m-го кадра обработанных исходного и синтезированного сигналов соответственно.

Мера сегментного отношения сигнал/шум [20]

$$d_{SEGSNR}(m)=\frac{10}{M}\sum_{m=0}^{M-1}\log\frac{\sum_{n=Nm}^{Nm+N-1}x^2(n)}{\sum_{n=Nm}^{Nm+N-1}\left(y(n)-x(n)\right)^2}, \tag{27}$$

где $n=\overline{0,N-1}$; N – число отсчетов в кадре сигнала; $x(n)$ и $y(n)$ – обработанные исходный и синтезированный сигналы текущего кадра. С учетом речевых пауз и вклада перцептуальных разностей сигналов нижняя и верхняя границы d_{SEGSNR} соответственно принимают значения –10 и +35 дБ.

Мера взвешенного спектрального наклона WSS [20] в дБ, основанная на использовании слуховой системы, в которой N перекрывающихся гауссовоподобных фильтров с увеличивающейся полосой частот используются для оценки сглаженного кратковременного спектра сигналов, задается соотношением

$$d_{WSS}(m)=\left(1\Big/\sum_{k=1}^{N}w(k)\right)\sum_{k=1}^{N}w(k)\left(E'_x(k)-E'_y(k)\right)^2, \tag{28}$$

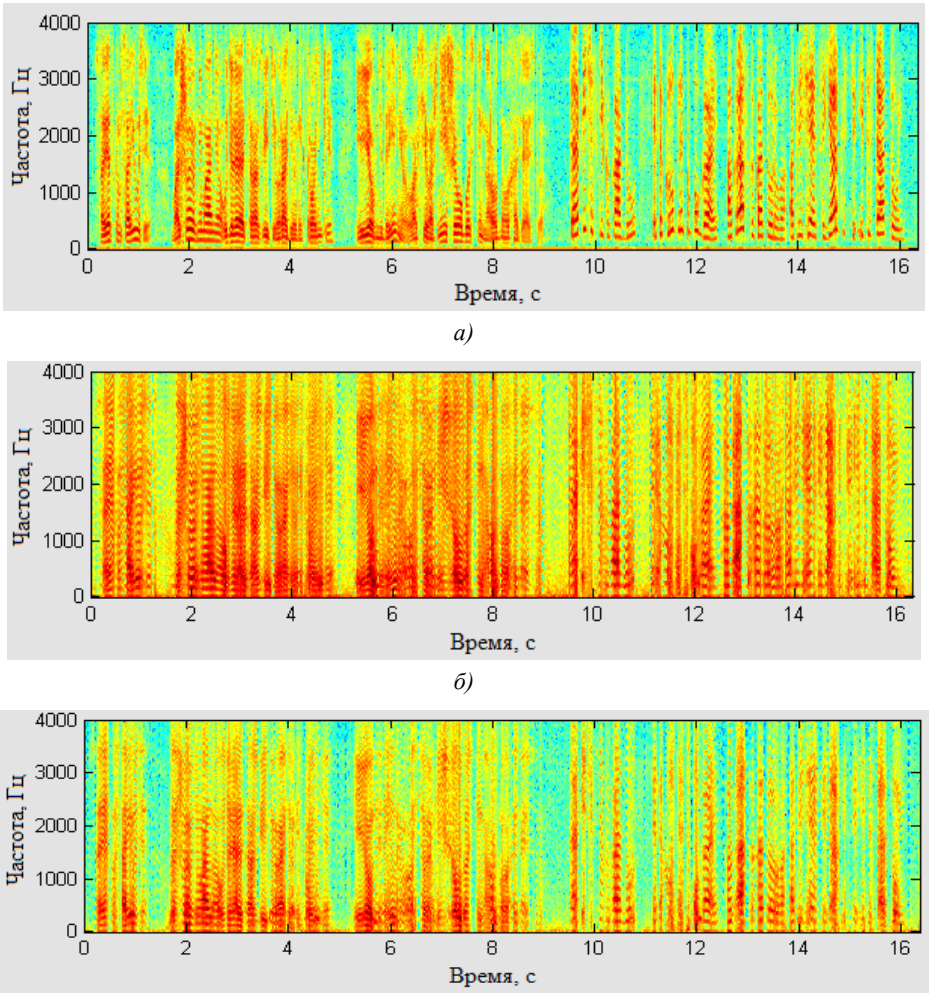
где $N=25$ – число критических частотных полос слуховой системы восприятия для определенного частотного диапазона [21]; $E'_x(k)=E_x(k+1)-E_x(k)$ и $E'_y(k)=E_y(k+1)-E_y(k)$ – наклоны первого порядка LPC-спектров мощности обработанных исходного и синтезированных сигналов $Y(k)$ в k-й полосе соответственно; $w(k)=\left(w_x(n)+w_y(n)\right)\big/2$ – весовой коэффициент для k-й полосы; $E_x(k)=10\lg\left(\max\left(S_x(n)\right)\right)$ и $E_y(k)=10\lg\left(\max\left(S_y(n)\right)\right)$ – максимальные значения выходных энергий фильтра k-й полосы для обработанных исходного и синтезированных сигналов; $S_x(n)$ и $S_y(n)$ – значения энергии n-й составляющей для каждого из фильтров; $w_j(k)=w_{j,\max}(k)w_{j,loc\max}(k)$; $w_{j,\max}(k)=\left(K_{j,\max}\Big/\left(K_{j,\max}+dB_{j,\max}-E_j(k)\right)\right)$ и $w_{j,loc\max}(k)=\left(K_{j,loc\max}\Big/\left(K_{j,loc\max}+dB_{j,loc\max}(k)-E_j(k)\right)\right)$ – весовые функции; j – индекс исходного или синтезированного сигналов; $K_{\max}=20$ и $K_{loc\max}=1$ – эмпирические значения, которые определяют коэффициент усиления для глобального и локального

максимумов; $dB_{\max}$ и $dB_{loc\max}(k)$ – значения глобального и локального максимумов для k-й частотной полосы.

Мера WSS использует весовые разности между спектральными наклонами в каждой критической полосе, так как спектральные вариации играют важную роль в восприятии качества речи, и отражает вероятное восприятие разборчивости.

Объективные количественные меры d_{IS}, d_{LAR}, d_{LLR}, d_{SEGSNR} и d_{WWS} вычислялись для каждого кадра длиной 240 отсчетов с использованием 75%-го перекрытия соседних кадров, взвешивания Hanning функцией [15], частоты дискретизации 8000 кГц и 10 LPC-коэффициентов.

Из сравнительного анализа спектрограмм исходного РС, состоящего из записей мужского и женского голосов, и РС, переданного с битовыми скоростями 290 и 390 бит/с и восстановленного с помощью векторного квантования с расщеплением (рисунок), видно, что с увеличением битового потока улучшается качество спектрограммы синтезированного РС за счет лучшей частотно-временной локализации его энергии.



а)

б)

в)

Спектрограммы сигналов в частотно-временной области при использовании частоты дискретизации 8 кГц и разрешения 8 бит на отсчет: *а)* исходный РС, переданный с битовой скоростью 64 кбит/с; *б)* восстановленный РС, переданный с битовой скоростью 290 бит/с; *в)* восстановленный РС, переданный с битовой скоростью 390 бит/с

В табл. 1 и 2 приведено распределение бит на параметры низкоскоростного кодека с битовой скоростью 290 и 390 бит/с, реализованного при помощи векторного квантования с расщеплением.

Таблица 1

Параметры кодека с битовой скоростью 290 бит/с на основе векторного квантования с расщеплением

Параметры суперкадра размером 800 отсчетов	Распределение бит/кадр			Число кадров в суперкадре	Бит/ суперкадр
LSF-коэффициенты	(1–3) LSF	(4–6) LSF	(7–10) LSF	1	17
	6	6	5		
Коэффициент усиления	2			3	6
Период основного тона	2			3	6
Общее число бит	21			–	29
Битовая скорость, бит/с	290				

Таблица 2

Параметры кодека с битовой скоростью 390 бит/с на основе векторного квантования с расщеплением

Параметры суперкадра размером 800 отсчетов	Распределение бит/кадр			Число кадров в суперкадре	Бит/ суперкадр
LSF-коэффициенты	(1-3) LSF	(4-6) LSF	(7-10) LSF	1	23
	7	7	9		
Коэффициент усиления	2			4	8
Период основного тона	2			4	8
Общее число бит	27			–	39
Битовая скорость, бит/с	390				

Уменьшение длительности Gain-кадра вызывает улучшение качества синтезированной речи из-за учета кратковременной динамики изменения звуковых объектов в РС и лучшей передачи низкоуровневых согласных звуков (табл. 3).

Таблица 3

Оценка влияния числа Gain-кадров в суперкадре на качество синтезированной речи для кодека с битовой скоростью 496 бит/с на основе векторного квантования с расщеплением

Мера искажений	Число Gain-параметров				
	4	5	6	7	8
IS	1,47	2,0	2,32	2,59	1,48
LAR	4,9	5,42	4,2	4,4	4,95
LLR	0,65	0,8	0,48	0,53	0,66
WSS	60,1	60,18	55,6	54,9	57,89
SEGSNR	0,16	0,15	0,135	0,14	0,18

Заключение

В настоящей статье разработан алгоритм низкоскоростного сжатия речи, основанный на использовании параметрической модели синтеза РС с линейным предсказанием, суперкадров с переменной структурой, векторного квантования параметров суперкадра и интерполяции LSF-кадров. Применение векторного квантования речевых параметров суперкадров уменьшает следующие виды избыточности: корреляцию между векторами, которые попадают в одну кодовую область, и корреляцию между компонентами вектора. Данный алгоритм обеспечивает возможность гибкого управления соотношением сжатие/качество синтезируемой речи и достижения компромисса между частотным и временным разрешениями РС за счет использования кадров различной длины для определения речевых параметров.

Приведены спектрограммы исходного и синтезированного сигналов, состоящих из записей мужских и женских голосов, для кодека с битовой скоростью 290 и 390 бит/с на основе векторного квантования с расщеплением. Установлено, что с увеличением битового потока улучшается качество спектрограммы синтезированного РС из-за лучшей частотно-временной локализации его энергии. Приведен сравнительный анализ влияния числа Gain-кадров в суперкадре

на качество синтезированной речи для кодека с битовой скоростью 496 бит/с на основе векторного квантования с расщеплением. Установлено, что увеличение битовой скорости обычно не приводит к значительному улучшению качества звучания из-за ограничений, накладываемых выбранной моделью речеобразования.

Осуществлена программная реализация в среде моделирования Matlab алгоритма низкоскоростного параметрического сжатия речи. Данная программа имеет следующие возможности: обучение и сохранение кодовой книги, изменение временной структуры суперкадра, использование сигнала возбуждения с одним или двумя состояниями, предыскажающего фильтра и двух способов векторного квантования (с расщеплением и многоуровневого). Результаты экспериментов показывают, что разборчивость речи сохраняется при битовых скоростях 300–800 бит/с.

Список литературы

1. Максимов, М.И. Проектирование низкоскоростных речепреобразующих устройств для каналов с высоким процентом ошибок / М.И. Максимов, Н.А. Сидорова, О.В. Чернояров // Электросвязь. – 2008. – № 7. – С. 48–49.

2. MELP: The new federal standard at 2400 bits/s / L.M. Supplee [et al.] // IEEE International Conference on Acoustics, Speech, and Signal Processing. – Munich, 1997. – P. 1591–1594.

3. Compandent's MELPe-Enhanced Mixed-Excitation Linear Predictive Vocoder [Electronic resource]. – Mode of access : http://www.compandent.com/products_melpe.htm. – Date of access : 03.03.2009.

4. Chamberlain, M. A 600 bps MELP vocoder for use on HF channels / M. Chamberlain // IEEE Military Communications Conference, MILCOM-2001, Communications for Network-Centric Operations: Creating the Information Force. – USA, 2001. – Vol. 1. – P. 447–453.

5. New NATO STANAG narrow band voice coder at 600 bit/s / G. Guilmin [et al.] // IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2006. – France Toulouse, 2006. – Vol. 1. – P. 689–692.

6. Wang, T. A 1200/2400 bps coding suite based on MELP / T. Wang, K. Koishida, V. Cuperman // Proc. of IEEE Workshop on Speech Coding. – Tsukuba, Japan, 2002. – Vol. 1. – P. 122–126.

7. Padellini, M. Very low bit rate (VLBR) speech coding around 500 bit/sec / M. Padellini, F. Capman, G. Baudoin // 12th European Signal Processing Conference (EUSIPCO 2004). – Vienna, Austria, 2004. – P. 1669–1672.

8. DARPA ASE. Program [Electronic resource]. – Mode of access : http://www.darpa.mil/ato/solicit/ASE/index.htm. – Date of access : 03.03.2009.

9. Kritzinger, C. Low Bit Rate Speech Coding [Electronic resource]. – Mode of access : etd.sun.ac.za/jspui/bitstream/10019/89/1/KritzC.pdf. – Date of access : 03.03.2009.

10. Попов, О.Б. Цифровая обработка сигналов в трактах звукового вещания / О.Б. Попов, С.Г. Рихтер. – М. : Горячая линия – Телеком, 2007. – 341 с.

11. Фант, Г. Акустистическая теория речеобразования / Г. Фант; пер. с англ. Л.А. Варшавского, В.И. Медведева ; под ред. В.С. Григорьева. – М. : Наука, 1964. – 284 с.

12. Маркел, Дж.Д. Линейное предсказание речи / Дж.Д. Маркел, А.Х. Грэй ; пер с англ. ; под ред. Ю.Н. Прохорова, В.С. Звездина. – М. : Связь, 1980. – 308 с.

13. Kabal, P. The computation of Line Spectral Frequencies Using Chebyshev Polynomials / P. Kabal, R.P. Ramachanandran // IEEE Trans. Acoustics, Speech, Signal Processing. – 1986. – Vol. 34, № 6. – P. 1419–1426.

14. Рабинер, Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М. : Радио и связь, 1981. – 496 с.

15. Марпл-мл., С.Л. Цифровой спектральный анализ и его приложения / С.Л. Марпл-мл. ; пер. с англ. – М. : Мир, 1990. – 584 с.

16. Linde, Y. An Algorithm for Vector Quantizer Design / Y. Linde, A. Buzo, R. Gray // IEEE Transactions on Communications. – 1980. – Vol. 28, № 1. – P. 84–94.

17. Real time vector quantization of LSP parameters / B. Kovesi [et al.] // Speech communication. – 1999. – Vol. 29, № 1. – P. 39–47.

18. Paliwal, K.K. Quantization of LPC Parameters / K.K. Paliwal, B.S. Atal [Electronic resource]. – Mode of access : maxwell.me.gu.edu.au/spl/publications/papers/book_sc_kkp.pdf. – Date of access : 03.03.2009.

19. Paliwal, K.K. Efficient vector quantization of LPC parameters at 24 bits/frame [Electronic resource]. – Mode of access : maxwell.me.gu.edu.au/spl/publications/papers/icassp91_kkp_lpc.pdf. – Date of access : 03.03.2009.

20. Hansen, J.H.L. An effective quality evaluation protocol for speech enhancement algorithms / J.H.L. Hansen, B.L. Pellom [Electronic resource]. – Mode of access : http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.9149. – Date of access : 03.03.2009.

21. Zwicker, E. Psychoacoustics, Facts and Models / E. Zwicker, H. Fast. – N.Y. : Springer-Verlag, 1990. – 354 p.

Поступила 17.02.09

Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: anbor@bsuir.by

A.A. Boriskevich, A.O. Rak

VECTOR-PARAMETRIC LOW-RATE COMPRESSION OF THE SPEECH SIGNAL ON THE BASIS OF A VARIABLE STRUCTURE SUPERFRAME

An algorithm of vector-parametric low-rate speech signal compression based on the parametric speech model with the linear prediction, a variable structure superframe, the vector quantization of superframe parameters (gain, pitch, and LSF coefficients), and the LSF-frame interpolation is proposed. The modeling results suggest that the algorithm allows to achieve low rate speech compression in the range of 300–800 bps under retaining speech intelligibility and control of quality/rate ratio at the expense of choice of the different number of frame within the superframe in computing speech model parameters. It was demonstrated that increasing bit rate doesn't lead to significant improvement of the speech quality due to the constraints imposed by the selected speech production model.