

На правах рукописи

РАХХАЛ МАХМУД

МЕТОДЫ И АЛГОРИТМЫ ВЕКТОРНО-РАЗНОСТНОГО КОДИРОВАНИЯ
ЦИФРОВЫХ АУДИОСИГНАЛОВ

Специальность 05.13.01 – Системный анализ, управление
и обработка информации (информатика)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург

2003

Работа выполнена на кафедре «Интеллектуальные системы управления» в государственном образовательном учреждении высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Научный руководитель: доктор технических наук, профессор
Гагарин Юрий Иванович

Официальные оппоненты: доктор физико-математических наук,
профессор
Иванов Владимир Михайлович

кандидат технических наук,
старший научный сотрудник
Егоров Владимир Викторович

Ведущая организация: Санкт-Петербургский государственный
электротехнический университет
(ЛЭТИ им. Ульянова-Ленина)

Защита состоится «18» марта 2004 г. в «16» часов на заседании диссертационного совета Д212.229.18 в государственном образовательном учреждении высшего профессионального образования «Санкт-Петербургский государственный политехнический университет» по адресу: 195251, Санкт-Петербург, Политехническая ул., 29, корпус 9, аудитория 325.

С диссертацией можно ознакомиться в библиотеке государственного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный политехнический университет».

Автореферат разослан «10» февраля 2004 г.

Ученый секретарь
диссертационного совета Д212.229.18

В.Н. Шашихин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Методы сжатия звуковых сигналов играют важную роль в цифровых мультимедийных приложениях, при организации голосовой связи, в компьютерных системах обучения и во многих других применениях компьютерных средств для управления и обработки информации.

В мультимедийных приложениях компьютерные средства сжатия аудио и видео данных нашли применение по трем главным причинам:

- большой объем мультимедийных данных;
- сравнительно медленные сохраняющие устройства или программы, которые не могут проигрывать мультимедийные данные в реальном времени;
- недостаточно высокая пропускная способность вычислительной сети, чтобы осуществлять передачу в реальном времени.

Широкое применение методы сжатия аудио и видеоданных нашли в системах компьютерного обучения с озвученным сопровождением экранной информации. Особенно это касается проблемы создания системы открытого образования, обеспечивающей доступ к образовательным ресурсам посредством применения информационных образовательных технологий дистанционного обучения, реализуемых на вычислительных сетях.

Важную роль играют методы сжатия при передаче звуковых сигналов по цифровым каналам связи в вычислительных сетях.

Существующие методы сжатия звуковых сигналов, которые нашли наибольшее практическое применение можно классифицировать следующим образом:

- кодирование формы сигналов,
- подполосное кодирование,
- кодирование через ортогональные преобразования.

Кодирование формы сигналов подразделяется на вокодерные и разностные алгоритмы, которые в свою очередь могут быть использованы

совместно с алгоритмами с линейным, в том числе с адаптивным предсказанием и алгоритмы с векторным квантованием (CELP-алгоритмы).

К кодированию формы сигнала относится метод векторно-разностного кодирования речевых сигналов, математическими моделями которого служат векторно-разностные уравнения.

В ряде опубликованных работ приведены результаты экспериментальных исследований, согласно которым по помехоустойчивости, по коэффициенту сжатия и качеству воспроизведенной речи к лучшим относятся алгоритмы с адаптивным предсказанием и алгоритмы с векторным квантованием (CELP).

Методы ВР-кодирования можно рассматривать как один из способов обобщения алгоритмов с адаптивным предсказанием и CELP-алгоритмов.

Таким образом, создание высококачественных методов и алгоритмов средств компрессирования звуковых сигналов в системах управления и вычислительных сетях является актуальной научно-технической задачей.

Цель и задачи исследования. Целью исследования является разработка методов и алгоритмов векторно-разностного кодирования на основе моделей линейных многоканальных систем и быстрых алгоритмов построения автокорреляционных матриц и решения заданных через них систем линейных уравнений.

В соответствии с поставленной целью в диссертации решались следующие задачи:

1. разработка методов векторно-разностного кодирования на основе моделей многоканальных линейных систем;
2. разработка быстрых алгоритмов вычисления автокорреляционных функций цифровых аудиосигналов;
3. разработка быстрых алгоритмов классификации аудиосигналов в методах ВР-кодирования;
4. разработка программных моделей для экспериментальных исследований;

5. разработка практических применений методов и быстрых алгоритмов ВР-кодирования.

Предмет исследований. Предметом исследований в диссертационной работе являются методы векторно-разностного кодирования звуковых сигналов с многоканальным линейным предсказанием, быстрые алгоритмы вычисления автокорреляционных функций, свертки и решения систем линейных уравнений, заданных через теплицеву автокорреляционную матрицу.

Научная новизна. В результате выполненных исследований предложены методы векторно-разностного кодирования с многоканальным линейным предсказанием, позволяющие при кодировании формы сигнала учитывать внутри и межвекторную корреляцию отсчетов звукового сигнала.

Предложены быстрые алгоритмы построения автокорреляционных матриц на основе функций средних значений разностей, позволяющих исключить операцию умножения вещественных чисел.

Предложены быстрые алгоритмы определения величины периода основного тона речевых сигналов на основе функций среднего значения разностей с прореженным скользящим окном.

Предложенные быстрые алгоритмы вычисления АКФ и классификации позволят создавать новые алгоритмы ВР-кодирования и повысить быстродействие известных алгоритмов сжатия аудиосигналов с линейным предсказанием.

Практическая ценность результатов исследований заключается в том, что предложенные в работе математические модели и алгоритмы апробированные с помощью компьютерного моделирования на различных реальных аудиосигналах, позволяют по сравнению с известными алгоритмами при одинаково высоком качестве воспроизведения сигнала и величины его компрессии повысить быстродействие декодирования.

Основные научные положения и результаты, выносимые на защиту :

1. Методы и быстрые алгоритмы ВР-кодирования с многоканальным линейным предсказанием.

2. Математические модели алгоритмов для вычисления автокорреляционных функций через быстрое преобразование Хартли и функции среднего значения разности.

3. Математические и программные модели для быстрой классификации типов и определения параметров звуковых сигналов.

Апробация работы. По результатам диссертационной работы опубликована одна статья в сборнике трудов СПбГПУ и два доклада в сборнике трудов Всероссийской конференции.

Внедрение результатов.

1. Комплекс программ моделирования ВР-методов сжатия аудиосигналов (СПбГПУ).
2. Информационно-справочная система с речевым сопровождением (Туристическая фирма).
3. Компьютерная обучающая система с речевым сопровождением (СПбГПУ, СПбГЭТУ).

Публикации. Основные результаты диссертационной работы опубликованы в 3 печатных работах, список которых приведен в конце автореферата.

Структура и объем работы. Диссертация состоит из введения, пяти глав, заключения, списка литературы и приложений. Работа содержит 126 страниц машинописного текста, 108 рисунков (из них 56 в приложениях) и список литературы на 46 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы научных исследований, приведены известная классификация общая оценка качества существующих методов сжатия цифровых аудиосигналов. Сформулирована цель и перечислены задачи исследований, приведена структура работы.

В первом разделе диссертации проведен анализ наиболее известных лучших методов сжатия, построенных на основе кодирования формы цифровых сигналов, включая методы векторно-матричного кодирования. В разделе приведены графики помехоустойчивости и оценок качества восстановленных речевых сигналов для различных методов сжатия аудиосигналов. Наибольшее внимание уделено анализу методов кодирования с кратко и долгосрочным предсказанием и с векторным квантованием.

Проведенный анализ позволил заключить, что для известных методов сжатия, которые имеют лучшие экспериментальные характеристики (кодирование с кратко и долгосрочным предсказанием и с векторным квантованием) отсутствуют общие математические модели, позволяющие дать на фундаментальной основе оценку их возможностей развития и алгоритмического практического воплощения.

Одним из подходов к совершенствованию математических форм предсказания указанных методов сжатия является метод векторно-разностного кодирования, в основе которого использовано векторное разностное уравнение

$$\bar{d}(n) = X(n) - \tilde{X}(n),$$

где $\bar{d}(n)$, $X(n)$, $\tilde{X}(n)$ - вектор в n -й дискретный момент времени соответственно погрешности предсказания, кодируемого цифрового сигнала, предсказанных значений кодируемого сигнала.

По своей сути векторно-разностное кодирование уже нельзя отнести к методу кодирования формы одномерного цифрового сигнала. Его следует рассматривать кодированием двумерного сигнала, образованного из одномерного посредством формирования векторов отсчетов с длиной, равной, например, периоду основного тона вокализованной речи (периодическими свойствами могут обладать и другие звуковые сигналы, например - музыкальные).

Векторно-разностное кодирование можно использовать в сочетании с алгоритмами с линейным предсказанием, в том числе по строкам и по

столбцам, если полагать, что в сформированном двумерном массиве строкам будут соответствовать одинаковые периоды основного тона.

Данная диссертационная работа посвящена проблемам развития и совершенствования математических моделей и алгоритмов ВР-кодирования во взаимосвязи с методами кодирования через векторное квантование и с адаптивным предсказанием.

Во втором разделе диссертации рассмотрены различные методы векторно-разностного кодирования: дельта-кодирование, ДИКМ-кодирование и кодирование с линейным многоканальным предсказанием. В начале раздела приведены графики цифровых сигналов, образованных последовательностями одноименных отсчетов, векторов $X(n)$ и графики последовательностей отсчетов внутри векторов. Длина векторов определялась периодом основного тона речевого сигнала. Формы представленных сигналов позволяют заключить, что имеется тесная корреляция между отсчетами как внутривекторных, так и межвекторных отсчетов. Поэтому в работе было акцентировано внимание на разработке методов ВР-кодирования на основе математических моделей многоканальных линейных систем, которые позволяют использовать как внутривекторную, так и межвекторную корреляцию отсчетов сигнала.

Применительно к принятым обозначениям в схемах ВР-кодирования и к параметрам цифрового фильтра модель многоканального линейного предсказания может быть записана в виде векторно-матричного уравнения

$$\tilde{X}(n) = \sum_{k=1}^p [A(k)] \hat{X}(n-k),$$

где $\tilde{X}(n)$ - вектор предсказанных значений для кодируемого вектора отсчетов сигнала $x(n)$, $[A(k)] = \|a_{ij}(k)\|$, $i, j = \overline{1, m}$ - матрица $m \times m$ коэффициентов $a_{ij}(k)$ линейного предсказания - коэффициентов цифрового фильтра, p - количество коэффициентов межвекторного предсказания

Разностному векторно-матричному уравнению соответствует структурная схема P -предсказателя ВР-кодирования (рис. 1), где в качестве P использован цифровой фильтр с импульсной характеристикой $[A(n-k)] = \|a_{ij}(n-k)\|$, $i, j = \overline{1, m}$,

$(\bar{a}_i(n-k), \hat{X}(n))$ - скалярное произведение вектора $\hat{X}(n)$ и i -ого вектора-строки $\bar{a}_i(n-k)$ матрицы $[A(n-k)]$,

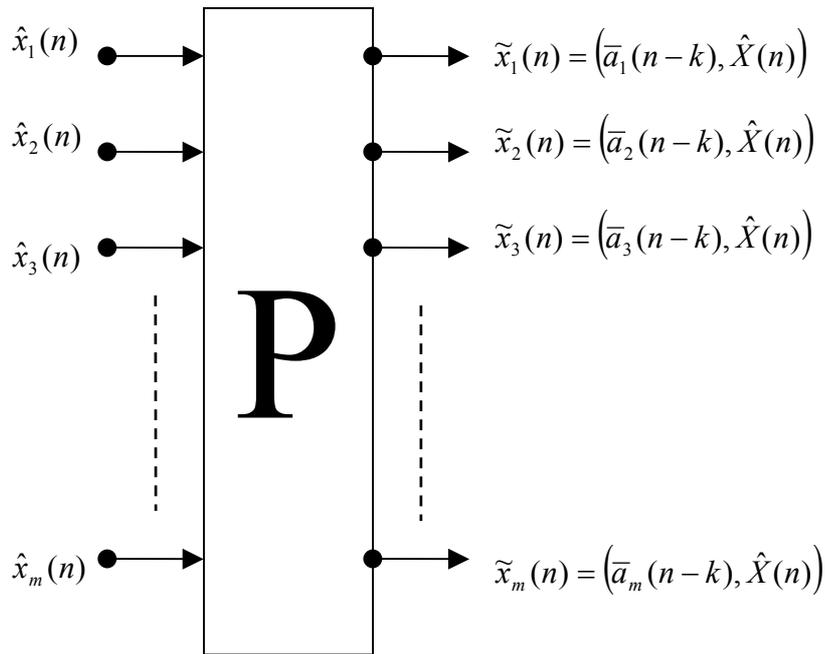


Рис. 1 Структурная схема предсказателя, представленная m -канальной линейной системой

В разделе приведены векторно-матричные соотношения для определения оптимальных коэффициентов a_{ij} -предсказания с использованием автокорреляционных функций сигналов. Приведены возможности использования известных методов быстрого решения систем линейных уравнений, заданных через автокорреляционные теплицевы матрицы. Показан пример построения быстрого алгоритма для определения квазиоптимальных коэффициентов предсказания.

В третьем разделе диссертации рассмотрены быстрые алгоритмы вычисления автокорреляционных функций, которые использованы в работе по трем назначениям:

- аperiodические АКФ для построения автокорреляционных матриц,
- аperiodические АКФ для определения величины периода основного тона,
- периодические АКФ для построения быстрого алгоритма вычисления аperiodических АКФ.

Быстрые алгоритмы вычисления АКФ на основе циклических матриц рассмотрены через дискретное преобразование Хартли (ДПХ). Левоциркулянтная матрица S_N связана с матрицей ДПХ следующим соотношением

$$S_N = H_N D_N^{(h)} H_N ,$$

где $D_N^{(h)} = P_N^{-1} D_N^{(F)} P_N$, $P_N = F_N H_N$ - переходная матрица от матрицы ДПФ к матрице ДПХ. Матрицу $D_N^{(h)}$ можно определить через коэффициенты $d_k^{(h)}$ ДПХ, полученные от образующего вектора-строки матрицы S_N

$$D_N^{(h)} = d_0^{(h)} \oplus 1/2 \left[\left(\bigoplus_{k=1}^{N-1} (d_{N-k}^{(h)} + d_k^{(h)}) \right) + \left(\bigoplus_{k=1}^{N-1} (d_{N-k}^{(h)} - d_k^{(h)}) \right) \bar{I}_{N-1} \right] ,$$

Различия формы для правого циркулянта, соответствующего периодической корреляции заключается в использовании матрицы инверсной перестановки строк. Так что выражение для правого циркулянта принимает вид:

$$\hat{S}_N = \bar{J}_N S_N = \bar{J}_N H_N [D_N^{(h)}] H_N ,$$

где $[D_N^{(h)}] = P_N^{-1} [D_N^{*(F)}] P_N$,

$\bar{J}_N = 1 \oplus \bar{I}_{N-1}$, \bar{I}_{N-1} - матрица инверсной перестановки,

В работе рассмотрены различные матрично-факторизованные формы матрицы H_N , которые соответствуют различным быстрым алгоритмам ДПХ (БПХ).

В работе рассмотрены также алгоритмы вычисления АКФ через функции среднего значения разностей (ФСР), которые для ограниченной выборки цифрового сигнала и с учетом оператора осреднения можно записать

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{m=0}^{N-1} |x(m)w_1(m) - x(m \pm k)w_2(m \pm k)|, \quad (1)$$

где N - величина скользящего окна,

$\pm k$ - величина сдвига окна (знак «+» или «-» указывает на направление смещения окна).

Для периодически продолженного сигнала ФСР аналогично периодической АКФ обладает симметрией

$$\hat{\gamma}(k) = \hat{\gamma}(N - k), \quad k = \overline{1, N-1},$$

или иначе - периодическая ФСР является четной функцией

$$\hat{\gamma}(k) = \hat{\gamma}(-k).$$

В случае использования прямоугольных весовых функций с единичной амплитудой для периодических ФСР выражение (1) принимает вид

$$\hat{\gamma}(k) = \frac{1}{N} \sum_{m=0}^{N-1} |x(m) - x(m+k)|,$$

В связи с этим появляется возможность вычисления смещенных оценок аperiodических АКФ через оценки ФСР, если исходное множество цифрового сигнала $\{x(m)\}_{m=0}^{N/2-1}$ дополнить $N/2$ нулями, т.е.

$$\{x(m)\}_{m=0}^{N-1} = \{x(0), x(1), \dots, x(N/2-1), 0, 0, \dots, 0\}.$$

Для сравнения оценок $\hat{R}(k)$ и $\hat{\gamma}(k)$ желательно использовать не прямые оценки $\hat{\gamma}(k)$, а функции $\tilde{\gamma}(k) = \hat{\gamma}_{\max} - \hat{\gamma}(k)$, т.е. дополнения $\hat{\gamma}(k)$ до $\hat{\gamma}_{\max}$.

Известно, что достоинством смещенных оценок АКФ является то, что построенные на их основе теплицевы автокорреляционные матрицы всегда

будут положительно полуопределенными, что в свою очередь гарантирует наличие решений системы линейных уравнений, заданных через данную матрицу.

Компьютерное моделирование показало (рис. 2), что для наиболее употребимых длин АКФ $N = 8 \div 16$ (рис. 2 в, г) в алгоритмах с линейным предсказанием успешно могут быть использованы для их вычисления функции $\tilde{\gamma}(k)$.

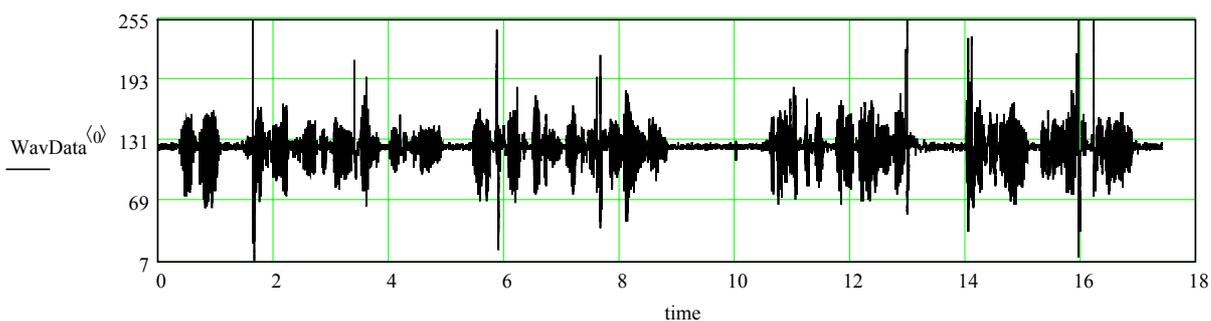


Рис. 2.(а) Пример речевого сигнала

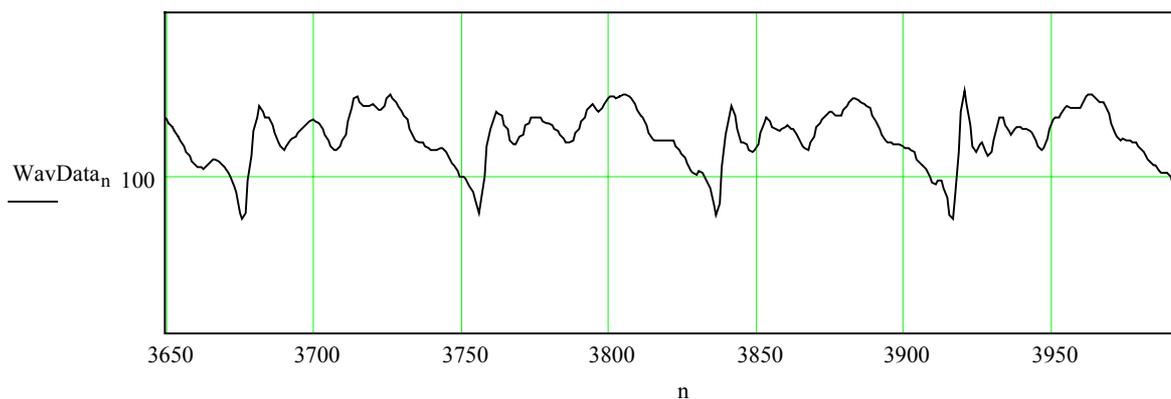


Рис. 2.(б) Фрагмент вышеприведенного речевого сигнала

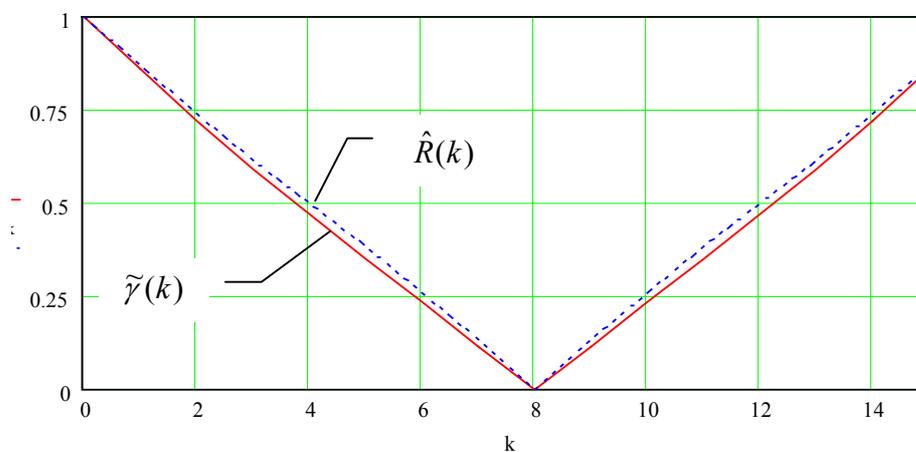


Рис. 2.(в) Графики оценок АКФ $\hat{R}(k)$ и ФСР $\tilde{\gamma}(k)$ для $N = 8$

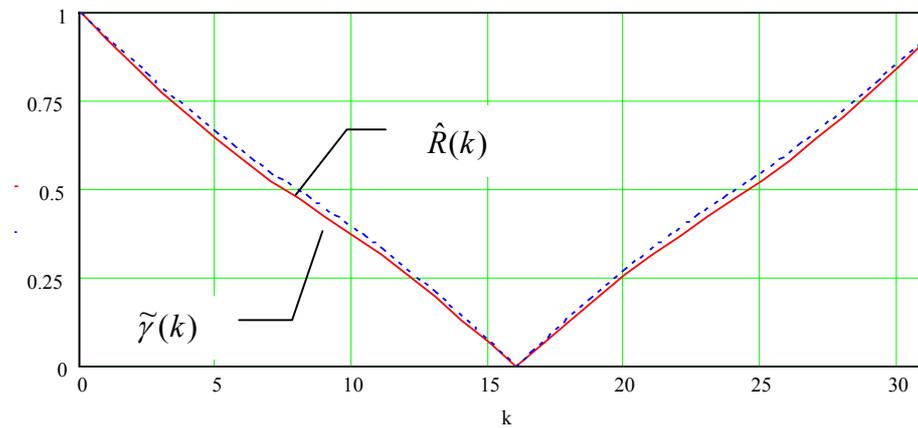


Рис. 2.(г) Графики оценок АКФ $\hat{R}(k)$ и ФСР $\tilde{\gamma}(k)$ для $N = 16$

В четвертом разделе рассмотрены математические модели и алгоритмы классификации и определения параметров звуковых сигналов.

Представлены способы классификации пауз, вокализованных и фрикативных фрагментов звуковых сигналов. Большое внимание уделено построению быстрых алгоритмов определения величины периода основного тона T_0 вокализованных фрагментов сигнала, на основе АКФ и ФСР.

Определение периода T_0 осуществлялось на исходном необработанном сигнале $x(n)$, на центрально-ограниченном $S\{x(n)\}$ и на трехуровневом центрально-ограниченном сигнале. В работе на реальных сигналах с помощью компьютерных моделей показано, что наибольшее предпочтение для определения T_0 необходимо отдать ФСР как по точности оценки, так и по количеству арифметических операций. Дальнейшего совершенствования алгоритмов вычисления АКФ и ФСР можно достичь за счет прореживания скользящего окна.

Формально процедуру прореживания можно выразить через разреженную прямоугольную весовую функцию

$$w(m) = \begin{cases} 1 & \text{при } n = 2m; \\ 0 & \text{при } n = 2m + 1, \end{cases}$$

$$n = \overline{0, N-1}, \quad m = \overline{0, \frac{N}{2}-1}.$$

В общем случае прореженную прямоугольную весовую функцию можно записать через Δ шаг приращения по индексу m внутри скользящего окна $\Delta = 1, 2, 3, \dots$

$$w(m) = \begin{cases} 1 & \text{при } m = 0, \Delta, 2\Delta, \dots, N-1 \\ 0 & \text{для всех других значений} \end{cases}$$

С учетом введенного окна выражения для АКФ и ФСР принимают вид

$$\hat{R}(k) = \frac{1}{M} \sum_{m=0}^{M-1} x(m) \cdot w(m) \cdot x(m+k), \quad (2)$$

$$(m+k) = \overline{0, N-1}, \quad k = \overline{0, (N_1 - N)}, \quad M = \frac{N}{\Delta}$$

$$\hat{\gamma}(k) = \frac{1}{M} \sum_{m=0}^{M-1} |x(m) \cdot w(m) - x(m+k)|. \quad (3)$$

Из выражений (2) и (3) следует, что уменьшения количества арифметических операций при использовании окна $w(m)$ достигается за счет уменьшения параметра осреднения.

Исследования влияния Δ - параметра на точность полученных оценок $\hat{R}(k)$ и $\hat{\gamma}(k)$ были выполнены с помощью компьютерного моделирования.

В пятом разделе приведен пример использования результатов исследования в виде компьютерной обучающей программы со звуковым сопровождением экранной информации.

ЗАКЛЮЧЕНИЕ

В процессе в диссертационной работе исследований получены следующие основные результаты:

1. Разработаны методы векторно-разностного кодирования с многоканальным линейным предсказанием, позволяющие использовать внутри и межвекторную корреляцию отсчетов кодируемого цифрового звукового сигнала.

2. Предлагаемые методы относятся к классу наиболее перспективных методов кодирования формы звуковых сигналов наряду с известными алгоритмами векторного квантования (CELP) и алгоритмами с долго и краткосрочным предсказанием. В отличие от указанных алгоритмов для ВР-кодирования разработанные математические модели на основе многоканальных линейных систем в форме векторно-матричных разностных уравнений.

3. Разработаны быстрые алгоритмы вычисления АКФ через БПХ и ФСР. Компьютерными моделями показано, что оценки предложенных дополненных до тах значения нормированные ФСР совпадают со смещенными оценками нормированных аperiodических АКФ, используемых для построения автокорреляционных матриц.

4. Разработаны математические и программные модели для алгоритмов определения величины периода основного тона вокализованных фрагментов звуковых сигналов через быстрые алгоритмы функций среднего значения разности, позволяющие сократить количество арифметических операций в несколько раз при сохранении требуемой точности оцениваемого периода.

5. Разработанные математические модели нашли практическое применение в компьютерных обучающих программах, например для анализа цифровых звуковых сигналов с помощью быстрых алгоритмов АКФ и ФСР.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Гагарин Ю.И., Раххал М. Математические модели предсказателей в схемах векторно-разностного кодирования речевых сигналов на основе многоканальных линейных систем. СПбГПУ серия ВТАРЭ № 487, С. 80.
2. Раххал М. Вычисление одномерных вещественных автокорреляционных функций через быстрое преобразование Хартли // Управление и информационные технологии: всероссийская научная конференция 3-4 апреля 2003 г. Санкт-Петербург. Сборник докладов в двух томах. Том. 2. 2003. – С. 202.
3. Раххал М. Вычисление автокорреляционных функций через функции среднего значения разности // Управление и информационные технологии: всероссийская научная конференция 3-4 апреля 2003 г. Санкт-Петербург. Сборник докладов в двух томах. Том. 2. 2003. – С. 204.