

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

1/2006

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

БЕСПРОВОДНАЯ СВЯЗЬ ●

БАНКИ ФИЛЬТРОВ ●

НЕЙРОСЕТЕВАЯ ОБРАБОТКА ●

ЦОС В РАДИОЛОКАЦИИ ●

ISSN 1684-2634



9 771684 263005 >

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 1/2006

Издается с 1999 года
Выходит четыре раза в год

ГЛАВНЫЙ РЕДАКТОР
Ю.Б. ЗУБАРЕВ

ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА:
В.В. ВИТЯЗЕВ, В.П. ДВОРКОВИЧ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:
Б.А. Бабаян, Ю.И. Борисов,
С.А. Грибачев, Г.В. Зайцев, Р.В. Зубарев,
А.П. Кирпичников, М.И. Кривошеев,
Н.А. Кузнецов, М.С. Куприянов, А.А. Ланнэ,
В.Г. Мистюков, С.Л. Мишенков,
А.А.Петровский, Ю.Н. Прохоров,
А.Н. Соловьев, Ю.Г. Сосулин,
В.В. Шахгильдян, Ю.С. Шинаков

Адрес редакции:
107241 Москва, Щелковское ш., 23А
Тел.: (495) 290-9088
Факс: (495) 290-9085
E-mail: editor@dspa.ru
http://www.dspa.ru

Для писем:
129090 Москва, а/я 48.

Издатель:
КБ волоконно-оптических приборов
Ответственный редактор: С.А. Задворнов
Компьютерная верстка: Е.М. Панова
Дизайн: И.Е. Артюхина

Подписной индекс по каталогу
ОАО «Роспечать» – **82185**

Подписано в печать 27.07.06
Формат 60х90/8.
Гарнитура «Arial». Печать ризографическая.
Бумага офсетная. Печ.л. 6,5. Тираж 700 экз.

Заказ № 3171. Отпечатано в
ООО НПЦ «Информационные технологии»
Рязань, ул. Островского, д. 21/1
тел.: (4912) 98-69-84

Издание зарегистрировано в Министерстве
Российской Федерации по делам печати,
телерадиовещания и средств
массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-1488
от 14.01.2000

УЧРЕДИТЕЛИ:

ЗАО «Инструментальные системы»
ФГУП «НИИ радио»

В НОМЕРЕ:

М.Г. Бакулин, В.Б. Крейнделин, А.П. Шумов
**Повышение скорости передачи информации и
спектральной эффективности беспроводных
систем связи.....2**

Л.В. Новиков
**Модификация банка фильтров вейвлетных
преобразований.....13**

А.В. Меркушева
**Нейросетевой метод цифровой обработки
нестационарного сигнала (речи) в вейвлет-
области.....18**

Н.Н. Вилкова
**Выбор стратегии управления научно-
исследовательскими и опытно-
конструкторскими работами.....31**

А.А. Кузин, А.В. Миронов, А.Г. Рындык, С.Б. Сидоров,
В.Д. Ястребов
**Модернизация РЛС на основе современных
технологий цифровой обработки сигналов.....34**

Г.В. Зайцев
**Вычислительный макроконвейер с переменным
тактом работы.....38**

Д. В. Алёшин, А.Т. Мингазин
**Программа для расчета экстремальных
параметров цифровых и аналоговых фильтров
и ее применение.....45**

**8-я Международная научно-техническая
конференция и Выставка
«Цифровая обработка сигналов и её
применение – DSPA'2006».....50**

Подписной индекс по каталогу
ОАО «Роспечать» – **82185**

ПОВЫШЕНИЕ СКОРОСТИ ПЕРЕДАЧИ ИНФОРМАЦИИ И СПЕКТРАЛЬНОЙ ЭФФЕКТИВНОСТИ БЕСПРОВОДНЫХ СИСТЕМ СВЯЗИ

М.Г. Бакулин, В.Б. Крейнделин, А.П. Шумов

Введение

Желание повысить скорость передачи информации в системах связи было всегда. Однако, несмотря на растущий спрос на высокоскоростное обслуживание, сделать это, особенно в беспроводных системах подвижной связи, очень сложно. Сейчас wireless local area networks (WLANs) обеспечивают скорости передачи до 10 Мбит/с. Но даже увеличение скорости передачи до 50 Мбит/с оказывается недостаточным. Речь может идти о нескольких сотнях мегабит в секунду. При использовании традиционных технологий передачи и приема сигналов такое увеличение скорости передачи данных может потребовать чрезмерно высокой излучаемой мощности или слишком большой полосы частот, что не всегда выполнимо. Кроме того, расширение спектра сигнала влечет за собой повышение несущей частоты, что в отсутствие прямой видимости может привести к существенному снижению дальности связи. Поэтому для достижения постоянно растущих скоростей передачи необходимо развитие новых телекоммуникационных технологий. Теоретические исследования и полученные практические результаты последних лет показали очень высокую эффективность использования в системах связи многоантенных приемных и передающих структур, при котором между приемником и передатчиком устанавливается канал со многими входами и многими выходами (multiple-input multiple-output (MIMO) channel). MIMO каналы позволяют существенно расширить возможности систем подвижной связи. Улучшение характеристик систем связи достигается благодаря тому, что MIMO каналы могут обеспечить *array gain* - выигрыш расстановки, *diversity gain* - выигрыш разнесения, *spatial multiplexing gain* - выигрыш пространственного мультиплексирования, *coding gain* - кодовый выигрыш, *interference reduction* - снижение уровня интерференции.

В работах [1, 2] рассмотрены вопросы пространственно-временного блочного и пространственно-временного решетчатого кодирования для многоантенных систем связи [3, 4].

Названные методы являются чрезвычайно важными, так как они позволяют резко повысить на-

Рассматриваются вопросы повышения скорости передачи информации в беспроводных системах связи. Приводятся результаты анализа пропускной способности МИМО каналов формулируются условия при которых возможно резкое повышение скорости передачи информации. Рассматриваются алгоритмы приема пространственно мультиплексированных сигналов а также пути реализации выигрыша пространственного мультиплексирования.

дежность передачи информации в условиях замираний. Пространственно-временные блочные коды дают высокое качество передачи информации за счет использования выигрыша разнесения при чрезвычайно низкой сложности декодирования. Пространственно-временные решетчатые коды обеспечивают еще более высокую надежность связи в условиях замираний за счет обеспечения выигрыша разнесения и кодового выигрыша, но при более высокой сложности декодирования. Оба вида пространственно-временного кодирования позволяют реализовать свои преимущества и при небольшом числе антенн на приемной стороне, что очень важно для мобильных станций.

Важнейшим ресурсом, позволяющим добиться высоких качественных показателей систем связи в условиях замираний, является разнесение и, в частности, пространственное разнесение. Оно обеспечивает наличие дополнительных путей для получения копий переданных сигналов. Очевидно, однако, что для повышения пропускной способности систем необходима организация параллельных каналов передачи разных потоков данных. В работах [5, 6] показано, что многоантенные системы обеспечивают не только повышенную надежность связи, но они также являются ключом к достижению значительно более высоких скоростей передачи информации без какого-либо расширения занимаемого диапазона частот.

В данной статье рассматриваются вопросы анализа пропускной способности многоантенных систем связи, а также пути реализации принципов, направленных на резкое повышение скорости передачи информации.

Пропускная способность МИМО каналов с общими релейскими замираниями

Общая схема МИМО системы приведена на рис. 1. Система имеет n_T передающих и n_R приемных антенн.

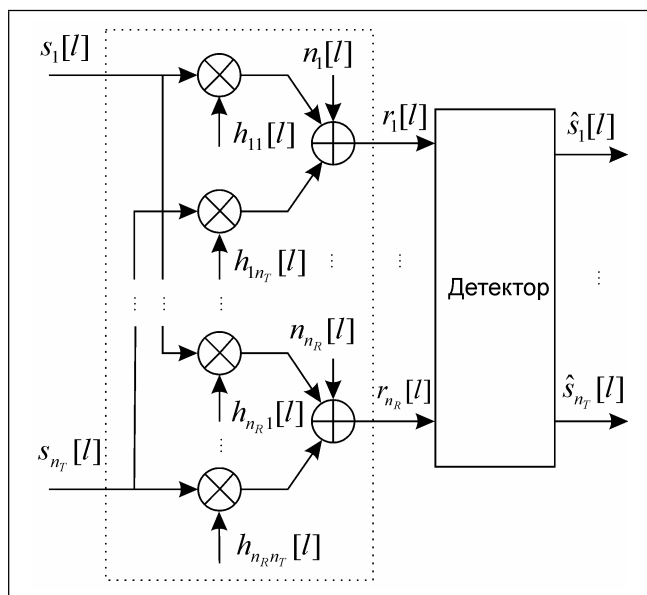


Рис.1. Блок-схема MIMO канала с n_T передающими и n_R приемными антеннами.

Будем считать замирания в MIMO канале релеевскими общими. Кроме того, предполагается, что канал остается постоянным в течение блока последовательных символов и меняется независимо от блока к блоку. Низкочастотный эквивалентный вектор принимаемых сигналов за один символичный период определяется соотношением

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

где $\mathbf{r} - n_R \times 1$ принятый сигнальный вектор, $\mathbf{s} - n_T \times 1$ переданный сигнальный вектор с $\langle \mathbf{s} \rangle = 0$, $\mathbf{H} - n_R \times n_T$ матрица канальных коэффициентов, $\mathbf{n}$ - аддитивный белый во времени комплексный гауссовский шум с $\langle \mathbf{n}\mathbf{n}^H \rangle = N_0 \mathbf{I}_{n_R}$. Полная передаваемая средняя мощность в символическом периоде считается ограниченной величиной P_S . В дальнейшем предполагается, что канал $\mathbf{H}$ точно известен (информация о состоянии канала на приемной стороне может поддерживаться через обучение и отслеживание). Применяя к матрице $\mathbf{H}$ сингулярное разложение - Singular Value Decomposition (SVD), можно записать следующие соотношения:

$$\mathbf{r} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^+ + \mathbf{n}$$

$$\mathbf{U}^+ \mathbf{r} = \mathbf{\Sigma}\mathbf{V}^+ \mathbf{s} + \mathbf{U}^+ \mathbf{n}$$

$$\tilde{\mathbf{r}} = \mathbf{\Sigma} \tilde{\mathbf{s}} + \tilde{\mathbf{n}}$$

где $\mathbf{U}$ и $\mathbf{V}$ - унитарные матрицы, такие, что распределения $\tilde{\mathbf{r}}$, $\tilde{\mathbf{s}}$ и $\tilde{\mathbf{n}}$ эквивалентны распределениям $\mathbf{r}$, $\mathbf{s}$ и $\mathbf{n}$ соответственно; знаком $^+$ отмечены псевдообратные матрицы Мура-Пенроуза; $\mathbf{\Sigma}$ - диагональная матрица, содержащая сингулярные величины $\mathbf{H}$, то есть положительные квадратные корни из собственных значений $\mathbf{H}\mathbf{H}^+$. Число ненулевых сингулярных величин определяет ранг матрицы $\mathbf{H}$. Для него справедливо соотношение $\text{rank} \mathbf{H} \leq \min(n_T, n_R)$. При этом MIMO канал развязывается на не более, чем $\min(n_T, n_R)$ параллельных комплексных подканалов с замираниями и белым шумом. Если элементы канальной матрицы $\mathbf{H}$ независимы и имеют одинаковые распределения, то с вероятностью 1 выполняется соотношение $\text{rank} \mathbf{H} = \min(n_T, n_R)$.

При заданной полной передаваемой мощности, ограниченной величиной P_S за символичный период, то есть

$$\sum_{j=1}^N |\tilde{s}_j|^2 = \text{tr} \langle \tilde{\mathbf{s}} \tilde{\mathbf{s}}^+ \rangle \leq P_S,$$

мгновенная пропускная способность MIMO канала может быть определена максимизацией взаимной информации [5]

$$I(\tilde{\mathbf{s}}; \tilde{\mathbf{r}}) = H(\tilde{\mathbf{r}}) - H(\tilde{\mathbf{n}}) = \log_2 \det \left(\frac{\mathbf{K}_{\tilde{\mathbf{r}}}}{N_0} \right),$$

где $H(x)$ - энтропия случайной переменной x , $\mathbf{K}_{\tilde{\mathbf{r}}} = \mathbf{K}_{\mathbf{r}}$ - корреляционная матрица принятого сигнального вектора:

$$\mathbf{K}_{\mathbf{r}} = \langle \mathbf{r}\mathbf{r}^+ \rangle = N_0 \mathbf{I}_{n_R} + \mathbf{H}\mathbf{K}_{\mathbf{s}}\mathbf{H}^+$$

$$\mathbf{K}_{\tilde{\mathbf{r}}} = N_0 \mathbf{I}_{n_R} + \mathbf{\Sigma}\mathbf{K}_{\tilde{\mathbf{s}}}\mathbf{\Sigma}^+.$$

Тогда пропускная способность определяется соотношениями

$$C_{|\mathbf{H}} = \max_{\text{tr} \mathbf{K} \leq \rho} \log_2 \det(\mathbf{I}_{n_R} + \mathbf{H}\mathbf{K}\mathbf{H}^+)$$

$$= \max_{\text{tr} \mathbf{K} \leq \rho} \log_2 \det(\mathbf{I}_{n_R} + \mathbf{\Sigma}\mathbf{K}\mathbf{\Sigma})$$

где, напомним, $\rho = P_S/N_0$ - отношение сигнал/шум в канале без замираний.

Эта пропускная способность является мгновенной, так как она обусловлена конкретной реализацией матрицы канальных коэффициентов $\mathbf{H}$. Когда информация о состоянии канала недоступна, пропускная способность достигается при комплексном гауссовском распределении с одинаковыми дисперсиями компонентов (circularly symmetric complex Gaussian distribution) сигнала с нулевыми средними значениями и корреляционной матрицей

$$\mathbf{K} = \frac{\rho}{n_T} \mathbf{I}_{n_T}.$$

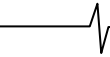
В этом случае выражение для мгновенной пропускной способности принимает вид

$$C_{|\mathbf{H}} \log_2 \det \left(\mathbf{I}_{n_R} + \frac{\rho}{n_T} \mathbf{H}\mathbf{H}^+ \right).$$

Теперь рассмотрим пропускную способность MIMO канала с замираниями при наличии точной канальной информации на приемной стороне.

Пропускная способность по Шеннону MIMO канала определяется наибольшей скоростью, на которой возможна безошибочная связь по всем возможным реализациям матрицы канальных коэффициентов передачи $\mathbf{H}$. Отметим, что для любой произвольно малой скорости передачи существует ненулевая вероятность того, что канальная матрица не может поддержать на этой скорости надежную связь. Это значит, что пропускная способность по Шеннону канала с релеевскими замираниями равна нулю. Учитывая это, для характеристики релеевских каналов используются другие теоретико-информационные характеристики, и, в частности, эргодическая пропускная способность и пропускная способность с отключениями [5, 7, 8].

Эргодическая пропускная способность определяется как усредненная пропускная способность по Шеннону по всем возможным реализациям $\mathbf{H}$



$$\bar{C}_{|\mathbf{H}} = \langle \log_2 \det(\mathbf{I}_{n_R} + \frac{\rho}{n_T} \mathbf{H}\mathbf{H}^+) \rangle_{\mathbf{H}}.$$

Отметим, что это усреднение часто выполняется более просто по $\mathbf{H}\mathbf{H}^+$, известному как Wishart матрица, когда элементы $\mathbf{H}$ являются независимыми релеевскими случайными переменными. Более подробно с этим вопросом можно ознакомиться по работам [5, 9] и ссылкам к ним.

В работах [5,10] было установлено, что при высоких SNR эргодическая пропускная способность определяется соотношением $C = \min(n_R, n_T) \log_2 \rho + O(1)$, которое ясно показывает линейный рост пропускной способности при увеличении наименьшего из чисел передающих и приемных антенн. На рис. 2 представлены

графики зависимости эргодической пропускной способности от отношения сигнал/шум для некоторых MIMO конфигураций. Как и ожидалось, эргодическая пропускная способность возрастает с возрастанием ρ , а также n_T и n_R . Отметим, что эргодическая пропускная способность канала с одним входом и многими выходами больше, чем эргодическая пропускная способность соответствующего канала со многими входами и одним выходом. Это связано с тем, что в отсутствие сведений о канале в передатчике каналы со многими входами и одним выходом не позволяют реализовать выигрыш расстановки (array gain.). Мы отсылаем заинтересованных читателей к [11,12,13] для анализа пропускной способности канала, когда ни передатчик, ни приемник не знают канальной матрицы $\mathbf{H}$.

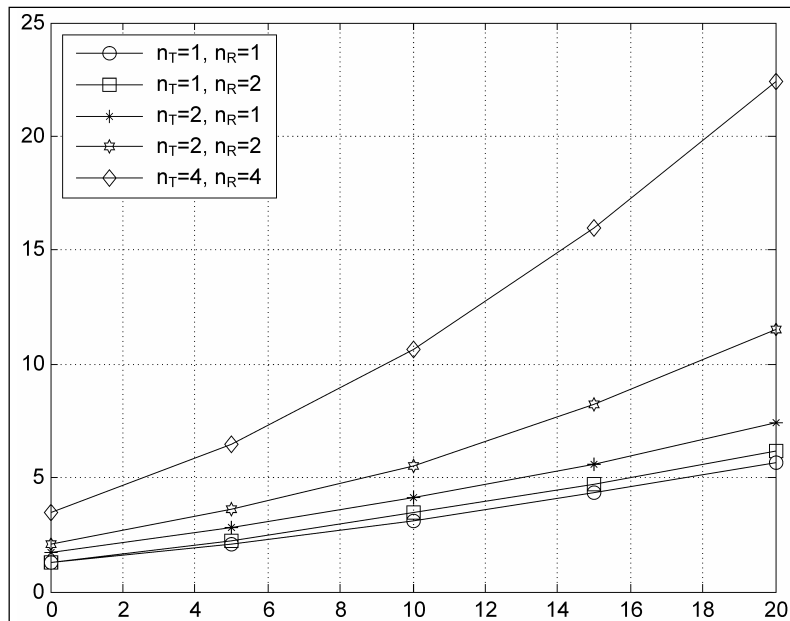


Рис. 2. Зависимость эргодической пропускной способности от отношения сигнал/шум для разных MIMO конфигураций.

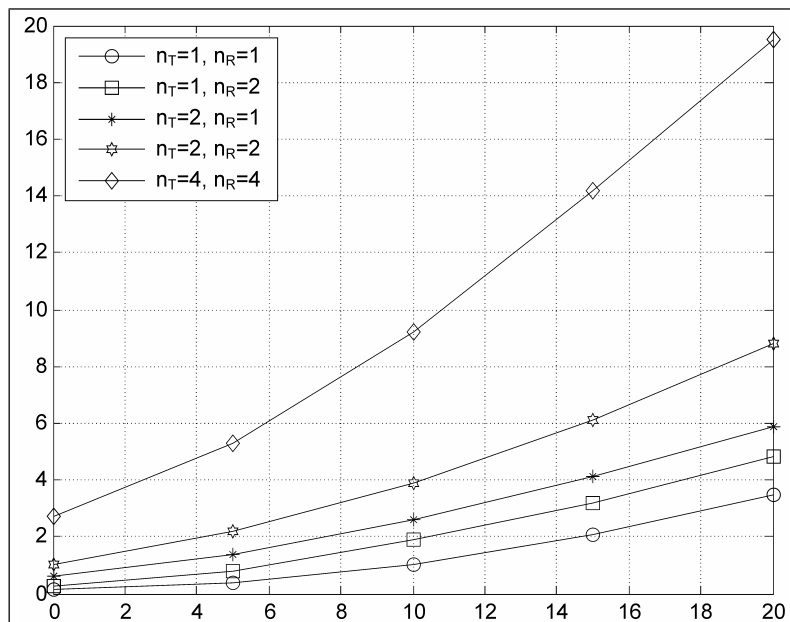


Рис. 3. Зависимость пропускной способности с 10% вероятностью отключения от отношения сигнал/шум для разных MIMO конфигураций.

Поскольку пропускная способность канала с релеевскими замираниями является случайной величиной, то она может быть охарактеризована не только своим средним значением (эргодической пропускной способностью), но и другими статистическими характеристиками. В качестве одной из таких характеристик используется функция распределения вероятности того, что мгновенная пропускная способность $C_{\text{н}}$ при конкретной реализации H меньше значения R скорости передачи информации $C_{\text{н}} < R$, то есть

$$F_C(R; \rho) = P(C_{\text{н}} < R).$$

Здесь $\rho = P_S / N_0$ - отношение сигнал/шум в канале без замираний, P_S есть полная мощность, излучаемая со всех антенн за символичный период. В дополнение к этому представлению используются также следующие количественные характеристики распределения.

Вероятность нарушения связи - outage probability. Для заданных скорости передачи R и излучаемой мощности, определяемой отношением сигнал/шум ρ , вероятность нарушения связи есть вероятность того, что пропускная способность канала недостаточна для поддержания надежной связи

$$P_{\text{OUT}}(R; \rho) = P(C_{\text{н}} < R) = F_C(R; \rho).$$

Дополнительная интегральная функция распределения - Complementary Cumulative Distribution Function (CCDF). Для заданного отношения сигнал/шум при передаче распределение (относительно скоростей передачи R) вероятности того, что пропускная способность канала достаточна для поддержания надежной связи

$$F_{\text{CCDF}}(R; \rho) = P(C_{\text{н}} \geq R) = 1 - F_C(R; \rho)$$

Поддерживаемая скорость – supportable rate. Для заданных максимальной вероятности нарушения связи ε и отношения сигнал/шум ρ , это есть максимальная скорость передачи, на которой надежная связь теоретически возможна

$$R_{\text{MAX}}(\varepsilon; \rho) = \max Rst(P(C_{\text{н}} < R) \leq \varepsilon) = F_C^{-1}(\varepsilon; \rho)$$

Этот параметр называется также **outage capacity** - пропускная способность с вероятностью нарушения связи не более ε .

Распределение пропускной способности и его числовые характеристики применительно к MIMO каналам сложны для аналитической оценки, кроме некоторых частных случаев.

На рис. 3 приведены графики зависимости 10% outage capacity - пропускной способности с 10% вероятностью отключения от отношения сигнал/шум для некоторых MIMO конфигураций [14]. Как и в случае эргодической пропускной способности, можно видеть, что пропускная способность с отключением возрастает с увеличением SNR, и что MIMO каналы дают значительное увеличение outage capacity. Фактически, поведение 10%-й outage capacity как функции SNR, n_T и n_R почти идентично поведению эргодической пропускной способности.

Итак, MIMO каналы могут дать беспрецедентное увеличение пропускной способности при увеличении числа

передающих и приемных антенн по сравнению с каналом с одной передающей и одной приемной антеннами. Однако утверждать, что всегда при использовании многочисленных антенн на передаче и на приеме произойдет увеличение пропускной способности нельзя. Например, в системах с STBC и STTC потоки сигналов для передачи с передающего антенного массива формируются так, что скорость передачи сигналов не может быть больше одного за такт, поэтому реализация резкого увеличения пропускной способности становится невозможной.

Рассмотреть влияние способа формирования потока сигналов для передачи по MIMO каналу на пропускную способность системы связи можно на примере системы связи с STBC. Соответствующий анализ проведен в работе [15]. По результатам анализа построены графики зависимости пропускной способности с отключением от отношения сигнал/шум для разного числа приемных и передающих антенн, приведенные на рис. 4, 5.

Из сопоставления графиков рис. 4 и рис. 5 видно, что использование STBC приводит к значительным потерям в пропускной способности MIMO каналов. Такой же вывод можно сделать и относительно использования пространственно-временных решетчатых кодов.

Таким образом, для резкого повышения скорости передачи информации в системах связи необходимо использовать алгоритмы передачи и приема сигналов, направленные на использование такого свойства MIMO каналов, как выигрыш пространственного мультиплексирования - spatial multiplexing gain.

Выигрыш пространственного мультиплексирования

MIMO канал может обеспечить линейный (в зависимости от минимального n_R или n_T) рост пропускной способности без дополнительной мощности или расширения спектра. Этот выигрыш в пропускной способности, называемый выигрышем пространственного мультиплексирования, реализуется передачей независимых сигналов с каждой из передающих антенн. При благоприятных канальных условиях, таких как сильное рассеяние, приемник может разделить разные потоки сигналов, обеспечивая линейный рост пропускной способности.

Приемники пространственно мультиплексированных сигналов

Общая схема MIMO системы приведена на рис. 1. С чисто математической точки зрения принятый сигнал в некоторый выборочный момент времени может быть выражен как

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n},$$

где $\mathbf{H}$ есть $n_R \times n_T$ матрица канальных коэффициентов. При полном ранге матрицы $\mathbf{H}$ разрешимость этого матричного уравнения определяется соотношением между n_R и n_T . До тех пор, пока $n_R \geq n_T$, переданный сигнальный вектор $\mathbf{s}$ может быть оценен. Если $n_R < n_T$, мы остаемся с недоопределенной системой, которая все же может быть частично разрешена, обычно при использовании некоторых дополнительных структур, налагаемых на множество передаваемых сигналов. Далее мы ограничим наше рассмотрение случаем $n_R \geq n_T$.

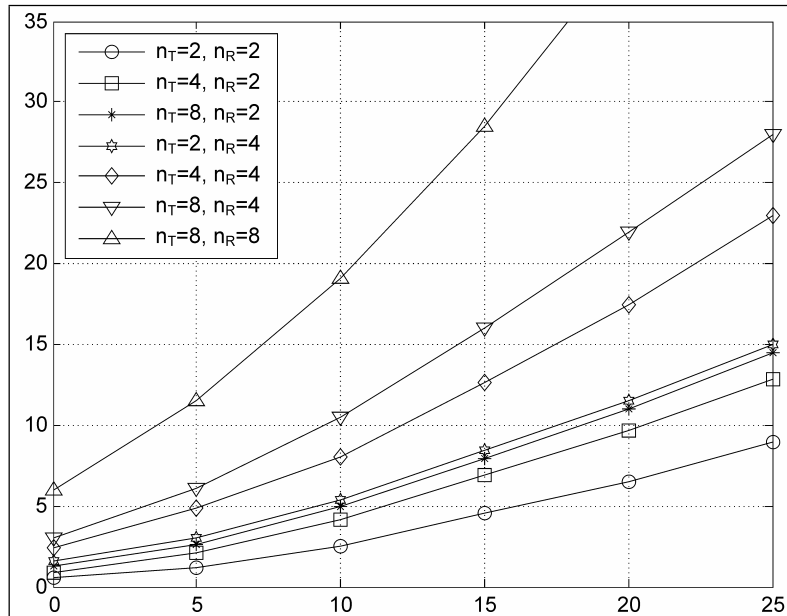


Рис. 4. Зависимость пропускной способности с 1% вероятностью отключения от отношения сигнал/шум для разных MIMO конфигураций.

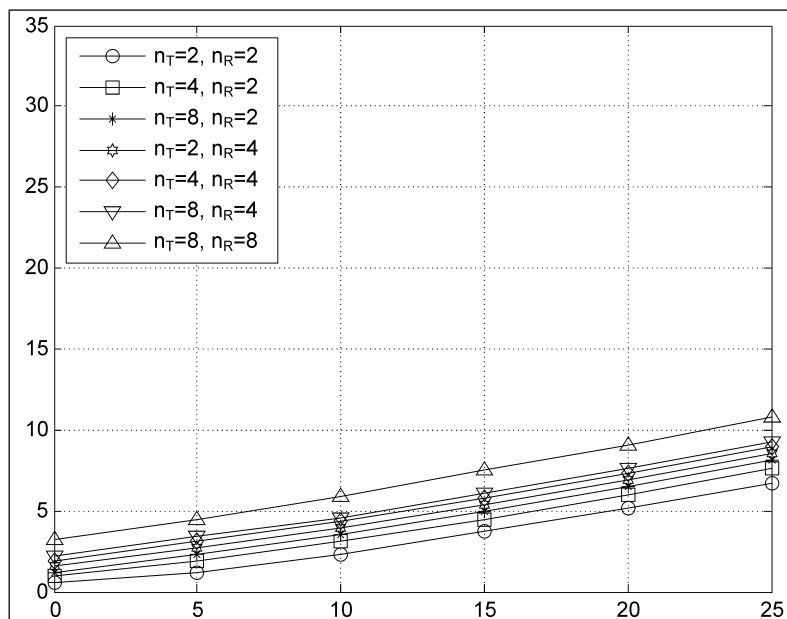


Рис. 5. Зависимость пропускной способности STBC с 1% вероятностью отключения от отношения сигнал/шум для разных MIMO конфигураций.

Проблема, стоящая перед приемником пространственно мультиплексированных сигналов, обусловлена присутствием межпоточной интерференции - multistream interference (MSI), поскольку сигналы, излученные с разных передающих антенн, взаимодействуют друг с другом и являются друг для друга помехами (напомним, что при пространственном мультиплексировании разные потоки данных передаются одновременно и, поэтому занимают один и тот же ресурс во времени и по частоте).

Проблема пространственного мультиплексирования лежит в создании эффективных с вычислительной точки зрения алгоритмов приемной обработки, которые позволяли бы в условиях

multistream interference (MSI) восстанавливать каждый переданный символ.

Перечислим существующие методы детектирования одновременно передаваемых сигналов.

Приемник максимального правдоподобия. Приемник максимального правдоподобия выполняет **векторное декодирование** и является оптимальным в смысле минимизации вероятности ошибки. При равновероятных и во времени некодированных векторных символах, МП приемник формирует оценку переданного сигнального вектора в соответствии с соотношением

$$\hat{\mathbf{s}} = \arg \min_s \|\mathbf{r} - \mathbf{H}\mathbf{s}\|^2,$$

где минимизация выполняется по всем возможным передававшимся векторам символом $\mathbf{s}$. Обозначим размер алфавита скалярного созвездия, передаваемого с каждой антенны через A . Прямая реализация МП алгоритма требует полного перебора по всем A^{n_T} векторным символам. Это делает сложность декодирования такого приемника exponential in the number - показательной от числа передающих антенн, и приемник становится практически не реализуемым. Однако, недавнее развитие быстрых алгоритмов [16-18] для методов сферического декодирования [19] дает возможность значительно понизить вычислительную сложность алгоритма (по меньшей мере для решетчатых кодов).

Линейные приемники. Сложность декодирования пространственно мультиплексированных сигналов можно существенно понизить за счет применения линейного фронтального приемника (linear receiver front-ends) (см. рис.6), чтобы сначала разделить переданные потоки данных, а затем независимо декодировать каждый из потоков.

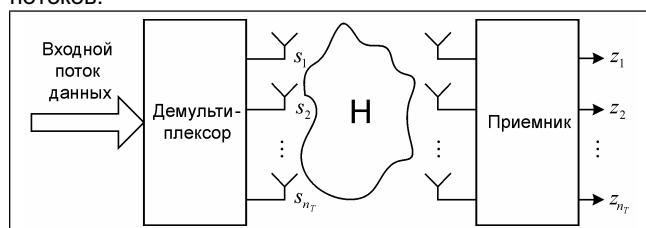


Рис.6. Структурная схема системы связи с линейным фронтальным приемником для разделения потоков данных, переданных через MIMO канал.

Процедура разделения символов определяется соотношением

$$\mathbf{z}(k) = \mathbf{W}(k)\mathbf{r}(k)$$

Под $\mathbf{W}(k)$ может пониматься:

- **Приемник с декорреляцией** (Zero-Forcing (ZF) receiver):

приемник с декорреляцией определяется соотношением

$$\mathbf{W}_{ZF}(k) = \mathbf{H}^+ = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H,$$

где $\mathbf{H}^+ = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$ означает псевдообращение Мура-Пенроуза канальной матрицы $\mathbf{H}$. Выход ZF приемника получается в виде

$$\mathbf{z} = \mathbf{s} + \mathbf{H}^+ \mathbf{n},$$

которое показывает, что ZF приемник разделяет матричный канал на n_T параллельных скалярных каналов с аддитивным пространственно-окрашенным шумом. Каждый скалярный канал затем декодируется независимо, игнорируя коррелированность шума в обрабатываемых потоках. ZF приемник преобразует проблему совместного декодирования в n_T проблем декодирования отдельных потоков (то есть он подавляет MSI), таким образом значительно снижая сложность приемника. Это снижение сложности происходит, однако, за счет увеличения шума, что приводит к значительному ухудшению качества по сравнению с МП приемником.

- Приемник, минимизирующий среднеквадратическую ошибку оценивания (Minimum Mean Square Error

(MMSE) receiver):

MMSE фронтальный приемник уравнивает уменьшение MSI и увеличение шума и определяется соотношением

$$\mathbf{W}_{MMSE}(k) = (\mathbf{H}^H \mathbf{H} + \frac{n_T}{\rho} \mathbf{I}_{n_T})^{-1} \mathbf{H}^H$$

В режиме низких SNR ($\rho \ll 1$) MMSE приемник приближается к согласованному фильтру, определяемому соотношением

$$\mathbf{W}_{MMSE} = N_0^{-1} \mathbf{H}^H$$

и превосходит ZF приемник, который увеличивает шум. На высоких SNR ($\rho \gg 1$)

$$\mathbf{W}_{MMSE} = \mathbf{W}_{ZF}.$$

Нелинейные приемники:

- Схемы, подобные схемам многопользовательского детектирования - Multi-User Detection (MUD) - нелинейные приемники от схем последовательного подавления интерференции до совместных максимально правдоподобных приемников;

- V-BLAST приемники: очень привлекательная архитектура, основанная на последовательном подавлении интерференции - Successive Interference Cancellation (SIC).

V-BLAST – архитектура, реализующая выигрыш пространственного мультиплексирования

В этом разделе рассматриваются основы построения беспроводной коммуникационной архитектуры, известной как Vertical-Bell Labs Layered Space Time Architecture или V-BLAST [20].

Система BLAST считается первой опубликованной системой, которая продемонстрировала реальные возможности MIMO каналов по существенному повышению попускной способности и спектральной эффективности. На лабораторном прототипе этой системы были продемонстрированы спектральные эффективности порядка 20-40 бит/с/Гц в условиях распространения внутри помещения на средних SNR от 24 до 34 дБ. Хотя эти результаты были получены в относительно благоприятных комнатных условиях, можно утверждать, что спектральные эффективности такого уровня беспрецедентны для беспроводных систем и просто недостижимы при использовании традиционных технологий.

Структура системы и основные принципы построения

Блок-схема системы BLAST приведена на рис.7.

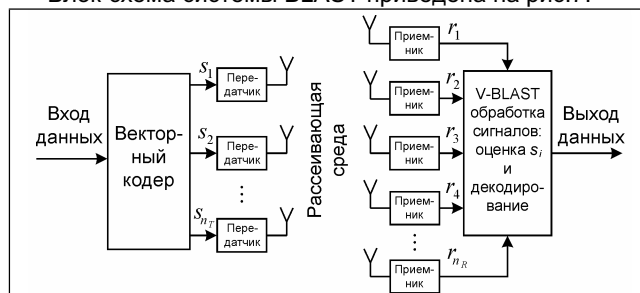
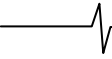


Рис. 7. Блок-схема системы V-BLAST с n_T передающими и n_R приемными антеннами.



Один поток данных демультиплексирован в n_T подпотоков, и каждый подпоток затем отображен в символы и направлен на соответствующий передатчик. Передатчики $1, \dots, n_T$ действуют одновременно на символьной скорости $1/T$ символов в секунду, с синхронизацией символов во времени. Каждый передатчик сам по себе является обычным QAM передатчиком. Набор передатчиков образует, в сущности, векторный передатчик, где компонентами каждого переданного n_T -мерного вектора являются символы, выбранные из QAM созвездия. Предполагается, что для каждого подпотока использовано одно и то же созвездие. Мощность, излучаемая каждым передатчиком, пропорциональна $1/n_T$, так что полная излучаемая мощность является постоянной и не зависит от n_T .

В V-BLAST передатчике выполняется операция демультиплексирования с последующим независимым бит-символьным отображением каждого подпотока. Никакого межподпоточного кодирования или кодирования какого-либо вида не требуется, хотя обычное кодирование индивидуальных подпотоков может быть применено. В данной статье полагается для простоты, что подпотоки содержат некодированные независимые символы данных.

Приемники $1, \dots, n_R$ есть отдельные обычные QAM приемники. Эти приемники также действуют совместно - каждый принимает сигналы, излученные со всех n_T передающих антенн. Для простоты в последующем предполагаются общие релейевские замирания, и матрица канальной передаточной функции - матрица $\mathbf{H}$ размерности $n_R \times n_T$, где каждый элемент h_{ij} есть комплексная передаточная функция от передатчика j к приемнику i и $n_T < n_R$. Предполагается, что изменение канала во времени на символьном периоде пренебрежимо мало, и что канал оценен точно, например, с использованием обучающей последовательности, так что для краткости мы не будем делать различий между $\mathbf{H}$ и ее оценкой.

Хотя V-BLAST есть по существу однопользовательская система, которая использует многочисленные передатчики, естественно спросить, чем подход BLAST отличается от использования традиционной техники множественного доступа в однопользовательском формате, то есть использования всех передатчиков для данных одного пользователя, которые были разбиты на подпотоки. Некоторые из этих отличий следует отметить. Первое. В отличие от кодового разделения или другого вида множественного доступа с расширением спектра, полоса частот используемая в BLAST системе, подобна диапазону, требуемому обычной системой с QAM. Второе. В отличие от FDMA каждый переданный сигнал занимает весь используемый диапазон. И последнее. В отличие от TDMA выделенный диапазон используется одновременно всеми передатчиками все время.

Взяты вместе эти отличия есть как раз то, что дает системе BLAST возможность реализовать более высокую спектральную эффективность, чем техники множественного доступа. Фактически, существенная черта BLAST состоит в том, что в ней нет явной ортогонализа-

ции переданных сигналов. Вместо этого сами условия распространения, которые, правда, предполагаются имеющими значительную многопутевость, используются для достижения **декоррелированности сигналов**, необходимой для их разделения. V-BLAST использует комбинацию старых и новых техник детектирования для эффективного разделения сигналов, что позволяет достичь значительной части Шенноновской пропускной способности и высокой спектральной эффективности.

Будем считать, что $\mathbf{s} = (s_1, s_2, \dots, s_{n_T})^T$ есть вектор передаваемых сигналов. Тогда соответствующий принятый n_R -мерный вектор есть

$$\mathbf{r}_1 = \mathbf{H}\mathbf{s} + \mathbf{n}$$

где $\mathbf{n}$ - вектор шума с компонентами, выбранными из независимых одинаково распределенных стационарных в широком смысле процессов с дисперсией σ^2 .

Алгоритм детектирования

Допустим, что упорядоченный набор

$$S \equiv \{k_1, k_2, \dots, k_{n_T}\}$$

является перестановкой целых чисел $1, 2, \dots, n_T$, устанавливающих порядок, в котором детектируются компоненты переданного символьного вектора $\mathbf{s}$. Процесс детектирования работает по $\mathbf{r}_1$, постепенно вычисляя решающие статистики $z_{k_1}, z_{k_2}, \dots, z_{k_{n_T}}$, которые затем квантуются для формирования оценок передававшихся символов $\hat{s}_{k_1}, \hat{s}_{k_2}, \dots, \hat{s}_{k_{n_T}}$. Таким образом, решающая статистика z_{k_1} вычисляется первой, затем вычисляется z_{k_2} и так далее. Позднее мы покажем, как определяется конкретный порядок S_{OPT} , оптимальный в некотором смысле, а сейчас рассмотрим общую процедуру детектирования относительно произвольной расстановки S_{OPT} .

Процесс детектирования использует линейное комбинационное обнуление и удаление символов для последовательного вычисления z_{k_i} и работает в общем следующим образом:

Шаг 1: Используя обнуляющий вектор $\mathbf{w}_{k_1}$, формируем линейную комбинацию из компонентов $\mathbf{r}_1$, чтобы получить решающую статистику z_{k_1} :

$$z_{k_1} = \mathbf{w}_{k_1}^T \mathbf{r}_1$$

Шаг 2: Квантуем z_{k_1} для получения $\hat{s}_{k_1}$:

$$\hat{s}_{k_1} = Q(z_{k_1})$$

где $Q(\cdot)$ означает операцию квантования (slicing), соответствующую используемому созвездию.

Шаг 3: Полагая, что $\hat{s}_{k_1} = s_{k_1}$, удаляем s_{k_1} из принятого вектора $\mathbf{r}_1$, получая модифицированный принятый вектор $\mathbf{r}_2$:

$$\mathbf{r}_2 = \mathbf{r}_1 - \hat{s}_{k_1} (\mathbf{H})_{k_1}$$

где $(\mathbf{H})_{k_1}$ означает k_1 -й столбец матрицы $\mathbf{H}$. Шаги 1-3 затем выполняются для компонентов $k_2, \dots, k_{n_T}$ в порядке последовательности модифицированных принятых векторов $\mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_{n_T}$.

Особенности процесса детектирования зависят от критерия, выбранного для вычисления обнуляющих векторов $\mathbf{w}_{k_i}$, наиболее общими из которых являются минимум среднеквадратической ошибки (MMSE) и zero-forcing (ZF). Процесс детектирования описан здесь применительно к ZF, поскольку это проще. k_i -й ZF обнуляющий вектор определяется как единственный с минимальной нормой вектор, удовлетворяющий условию

$$\mathbf{w}_{k_i}(\mathbf{H})_{k_j} = \begin{cases} 0 & j \geq i \\ 1 & j = i \end{cases}$$

Таким образом, $\mathbf{w}_{k_i}$ является ортогональным к подпространству, охваченному вкладами в $\mathbf{r}_i$ тех символов, которые не оценены, и не удалены. Можно показать, что единственный вектор, удовлетворяющий этому условию, есть как раз k_i -я строка $\mathbf{H}_{k_i}^+$, где обозначение $\mathbf{H}_{k_i}^+$ означает матрицу, полученную обнулением столбцов $k_1, k_2, \dots, k_i$ матрицы $\mathbf{H}$.

MMSE обнуление более детально обсуждается в литературе по адаптивным антенным решеткам, например в [21]. В любом случае, однако, мощность шума k_i -й решающей статистики z_{k_i} пропорциональна

$$\|\mathbf{w}_{k_i}\|^2,$$

и таким образом, последетекторные SNR пропорциональны

$$1/\|\mathbf{w}_{k_i}\|^2.$$

Полный ZF V-BLAST алгоритм детектирования может теперь быть описан как рекурсивная процедура, включающая определение оптимального упорядочивания, следующим образом:

инициализация: $\mathbf{G}_1 = \mathbf{H}^+$ (a)

$i = 1$ (b)

рекурсия: $k_i = \arg \min_{j \notin \{k_1, \dots, k_{i-1}\}} \|(G_i)_j\|^2$ (c)

$\mathbf{w}_{k_i} = (\mathbf{G}_i)_{k_i}$ (d)

$z_{k_i} = \mathbf{w}_{k_i}^T \mathbf{r}_i$ (e)

$\hat{s}_{k_i} = Q(z_{k_i})$ (f)

$\mathbf{r}_{i+1} = \mathbf{r}_i - \hat{s}_{k_i} \mathbf{w}_{k_i}$ (g)

$\mathbf{G}_{i+1} = \mathbf{H}_{k_i}^+$ (h)

$i = i + 1$ (i)

где $(\mathbf{G}_i)_j$ есть j -я строка $\mathbf{G}_i$. Таким образом, соотношение (c) определяет элементы S_{OPT} - оптимальный порядок, обсуждаемый позже. Операторы (d-f) вычисляют соответственно ZF обнуляющий вектор, решающую статистику и

тор, решающую статистику и оцененные компоненты $\mathbf{s}$. Уравнение (g) выполняют удаление продетектированных компонентов из принятого вектора и (h) вычисляет новую псевдоинверсию для следующей итерации. Заметим, что эта новая псевдоинверсия основана на версии $\mathbf{H}$ пониженного порядка, в которой столбцы $k_1, k_2, \dots, k_i$ были обнулены. Эти столбцы соответствуют компонентам $\mathbf{s}$, которые уже были оценены и удалены и, таким образом, система становится эквивалентной версии рис.1 пониженного порядка, в которой передатчики $k_1, k_2, \dots, k_i$ были удалены или эквивалентна системе, в которой $s_{k_1} = \dots = s_{k_i} = 0$.

Определение S_{OPT} . Напомним, что согласно принятому предположению все компоненты $\mathbf{s}$ принадлежат одному и тому же созвездию. При этом предположении z_{k_i} с наименьшим последетекторным SNR будут определять помехоустойчивость процесса детектирования. Важный аспект нелинейной обработки в этой схеме состоит в том, что из-за удаления символов эти последетекторные SNR зависят от порядка, в котором вычисляются решающие статистики. Таким образом, очевидное достоинство этой системы - максимизация последетекторного SNR, которое на данном шаге определяет надежность вычисляемой решающей статистики. Можно показать, что локальная оптимизация (уравнение (c)) выбора компонента с лучшим SNR на каждой стадии ведет, что отчасти удивительно, к глобальному оптимуму S_{OPT} .

Результаты моделирования

На рис. 8 приведены графики зависимости вероятности ошибки от отношения сигнал/шум для систем V-BLAST с четырьмя передающими и четырьмя приемными антеннами при использовании QPSK модуляции и при разных алгоритмах приема.

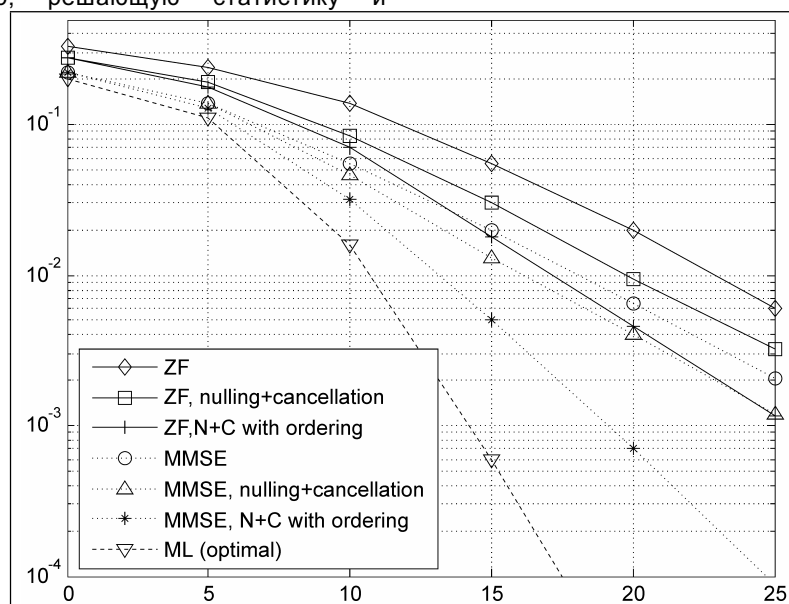
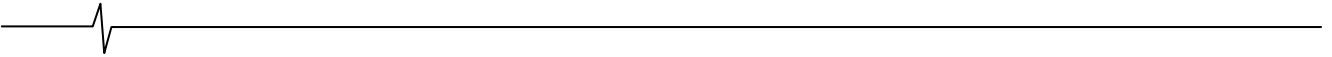


Рис. 8. Графики зависимостей вероятности ошибки от отношения сигнал/шум для системы V-BLAST с четырьмя передающими и четырьмя приемными антеннами и QPSK модуляцией при разных алгоритмах приема.



Из графиков видно, что ZF и MMSE приемники уступают по качеству как алгоритмам с удалением продетектированных символов и упорядочиванием, так и МП детектору. Несмотря на это, ZF и MMSE обнуляющие стратегии широко используются на практике из-за их существенно более низкой вычислительной сложности. Сравнивая квазиоптимальные методы между собой, можно видеть, что наименьшую вероятность ошибки обеспечивает MMSE прием с обнулением, удалением и упорядочиванием.

Относительно выигрыша пространственного мультиплексирования можно отметить, что система V-BLAST передает n_T символов за смвольный период, и достигает выигрыша пространственного мультиплексирования в $\min(n_T, n_R)$. Если $n_R > n_T$, все n_T символов могут быть восстановлены. В противном случае, не больше n_R символов могут быть надежно переданы.

Этот выигрыш мультиплексирования $\min(n_T, n_R)$ считается оптимальным, поскольку математически невозможно разделить более чем $\min(n_T, n_R)$ символов из n_R наблюдаемых смесей.

Результаты лабораторных испытаний

На рис. 9 приведены результаты, полученные с использованием лабораторного прототипа системы рис.1 с 8 передатчиками и 12 приемниками [20]. Система работала на несущей частоте 1,9 ГГц и символьной скорости 24.3 ксимв./с в диапазоне 30 кГц, используя некодированную 16-QAM в каждом передатчике при спектральной эффективности

$$\frac{(8 \text{ Пер}) \times (4 \text{ бит/симв/Пер})(24.3 \text{ ксимв/с})}{30 \text{ кГц}} = 25,9 \text{ бит/с/Гц}$$

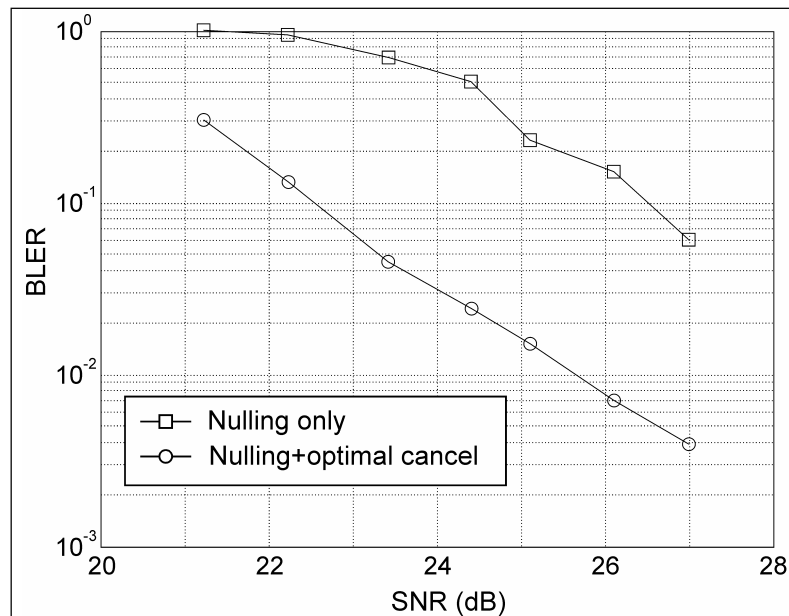


Рис. 9. Зависимости вероятности ошибки на блок от отношения сигнал/шум, полученные с использованием лабораторного прототипа системы рис.1 с 8 передатчиками и 12 приемниками (модуляция – 16-QAM).

Верхняя кривая на рис.9 показывает качество полученное, когда использовалось только обнуление. Нижняя кривая показывает качество при использовании обнуления и удаления с оптимальным упорядочиванием. Средняя разница составляет порядка 4 дБ, что соответствует приблизительно разнице 10 бит/с/Гц в спектральной эффективности.

В эксперименте символы передавались пакетами по 100 символов, 20 из которых использовались для оценки канала, так что полезная нагрузка составляла 80% от спектральной эффективности или 20,7 бит/с/Гц. При отношении сигнал/шум 34 дБ спектральная эффективность порядка 40 бит/с/Гц была продемонстрирована при тех же вероятностях ошибок. Все результаты были получены на небольших дальностях (порядка 12 м) внутри помещения при пренебрежимо малом разбросе задержек.

Результаты испытаний показали, что достигнутая

спектральная эффективность беспрецедентна для беспроводных каналов, и такую спектральную эффективность невозможно получить при традиционных подходах, в которых используется один передатчик, просто потому что требуемая нагрузка на созвездие будет чрезмерна. Например, для получения 32 бит на векторный символ, как в экспериментах, при использовании одного передатчика потребуется созвездие, состоящее из 2^{32} (или более, чем 10^9) точек, которые находятся далеко за пределами реальности независимо от SNR.

Сопоставление V-BLAST и STBC

Возросшая пропускная способность, обеспечиваемая схемой V-BLAST, не приходит без потерь в качестве. На рис. 10 сравниваются зависимости вероятности ошибок для систем с V-BLAST и с Alamouti STBC, работающих через один и тот же 2x2 MIMO канал. Чтобы сделать справедливое сравнение, скорость передачи

данных задана одинаковой. Скорость кода Аламаути может быть повышена использованием модуляции более высокого порядка. Таким образом, мы анализируем качество по результатам моделирования для STBC, используя QPSK, и для V-BLAST, используя BPSK, дающие в обеих схемах максимальную скорость в 2

бит/с/Гц. Замирания считаются релейевскими общими. Также для поддержания приемлемых уровней сложности приемника в системе V-BLAST использовался алгоритм MMSE с удалением и оптимальным упорядочиванием.

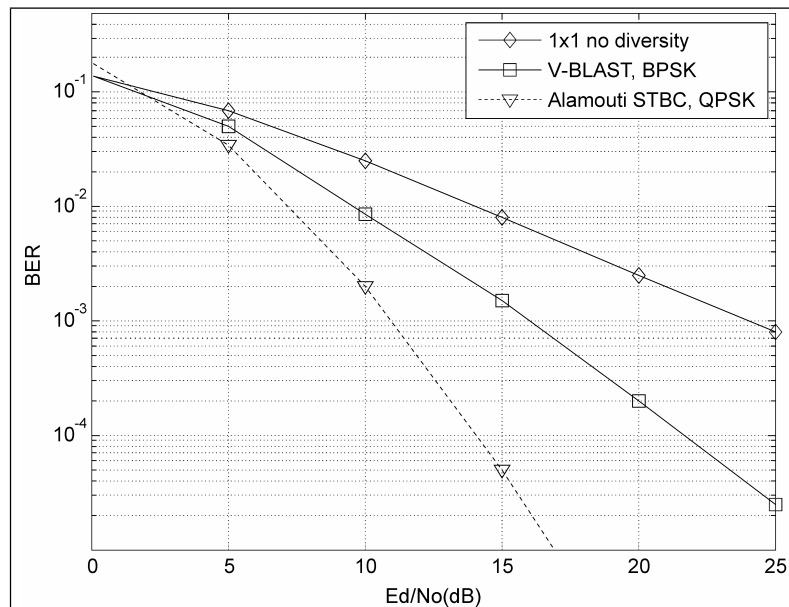


Рис. 10. Зависимость вероятности ошибки от отношения сигнал/шум для систем V-BLAST и Аламаути STBC с двумя передающими и двумя приемными антеннами при действии общих релейевских замираний.

Из сопоставления графиков видно, что блочный пространственно-временной код существенно превосходит V-BLAST на большинстве SNR, что не удивительно, так как STBC обеспечивает МП качество. Правда, при этом надо помнить, что скорости передачи для V-BLAST систем, которые тестировались в Bell Labs, составляют 20-40 бит/с/Гц. Этот порядок скоростей может быть легко достигнут, с использованием BLAST, но он недостижим для систем с STBC.

Заключение

Из приведенных результатов моделирования и испытаний прототипа системы V-BLAST видно, что система V-BLAST оправдывает связанные с ней ожидания. Она впервые продемонстрировала возможности MIMO каналов по резкому повышению скорости передачи и спектральной эффективности систем связи. Система V-BLAST демонстрирует беспрецедентную скорость передачи данных в условиях замираний и весьма эффективные алгоритмы детектирования, позволяющие реализовать multiplexing gain. Однако из результатов анализа также следует, что квазиоптимальные алгоритмы, имеющие высокую вычислительную эффективность, не в полной мере реализуют возможности MIMO систем по качеству детектирования. По уровню вероятности ошибки эффективные с вычислительной точки зрения квазиоптимальные алгоритмы уступают максимально правдоподобному детектированию, прямое применение которого при высокой скорости передачи, однако, невозможно из-за вычислительной сложности. Поэтому важной за-

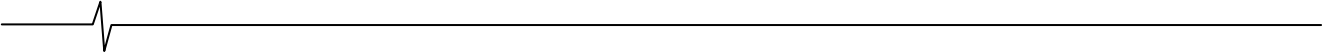
дачей является разработка алгоритмов детектирования с характеристиками МП, но с практически реализуемой сложностью. Среди других направлений, в которых ведутся интенсивные исследования необходимо назвать такие как

- использование пространственно-временного кодирования в системах высокоскоростной передачи информации;
- совместное использование помехоустойчивого канального и пространственно-временного кодирования в высокоскоростных системах;
- разработка приемлемых по сложности высокоскоростных алгоритмов передачи и приема сигналов в системах с числом передающих антенн, меньшим числа приемных антенн.

Совместно с другими исследованиями результаты, получаемые по этим направлениям, в недалекой перспективе позволят существенно повысить скорость и надежность передачи информации в системах подвижной связи.

Литература

1. А.П. Шумов, М.Г. Бакулин, В.Б. Крейнделин. Методы пространственно-временного блочного кодирования // Цифровая обработка сигналов.- 2005.- №2. С. 2-10.
2. А.П. Шумов, М.Г. Бакулин, В.Б. Крейнделин. Пространственно-временные решетчатые коды для многоантенных систем связи // Цифровая обработка сигналов.- 2005.- №3. С. 38-47.
3. V. Tarokh, H. Jafarkhani, and A. Calderbank, «Space-time

- 
- block codes from orthogonal designs,» *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.
4. V. Tarokh, N. Seshadri, and A. R. Calderbank, «Space–time codes for high data rate wireless communication: Performance criterion and code construction,» *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
 5. I. E. Telatar, «Capacity of multi-antenna Gaussian channels,» *Eur. Trans. Tel.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
 6. G. J. Foschini and M. J. Gans, «On limits of wireless communications in a fading environment when using multiple antennas,» *Wireless Personal Communications.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
 7. L. H. Ozarow, S. Shamai, and A. D. Wyner, «Information theoretic considerations for cellular mobile radio,» *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
 8. E. Biglieri, J. Proakis, and S. Shamai, «Fading channels: Information-theoretic and communications aspects,» *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
 9. A. Scaglione. Statistical analysis of the capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs. In International Symposium on Information Theory, page 278, June 2002.
 10. O. Oyman, R.U. Nabar, H. Bölcskei, and A.J. Paulraj, «Characterizing the statistical properties of mutual information in MIMO channels,» *IEEE Trans. Signal Processing*, vol. 51, pp. 2784–2795, Nov. 2003.
 11. T. L. Marzetta and B. M. Hochwald, «Capacity of a mobile multipleantenna communication link in Rayleigh flat fading,» *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
 12. B. M. Hochwald and T. L. Marzetta, «Unitary space–time modulation for multiple antenna communications in Rayleigh fading,» *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.
 13. L. Zheng, D. N. C. Tse, «Communicating on the Grassmann manifold: A geometric approach to the noncoherent multiple antenna channel,» *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.
 14. A.J. Paulraj, D.A. Gore, R.U. Nabar, H. Bolecskei. An Overview of MIMO Communications – A Key to Gigabit Wireless. Proceedings of IEEE, Vol.92, No.2, p. 198-218, February 2004.
 15. S. Sandhu and A. Paulraj. Space-time block codes a capacity perspective. IEEE Communications Letters, vol. 4, no. 12, p. 384 -386, December 2000.
 16. E. Viterbo and J. Buotros, «A universal lattice code decoder for fading channels,» *IEEE Trans. Inform. Theory*, vol. 45, pp.1639–1642, July 1999.
 17. O. Damen, A. Chkeif, and J. C. Belfiore, «Lattice code decoder for space–time codes,» *IEEE Commun. Lett.*, vol. 4, pp. 161–163, May 2000.
 18. B. Hassibi and H. Vikalo, «On the expected complexity of sphere decoding,» in *Proc. Asilomar Conf. Signals, Systems and Computers*, vol. 2, 2001, pp. 1051–1055.
 19. U. Fincke and M. Pohst, «Improved methods for calculating vectors of short length in a lattice, including a complexity analysis,» *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.
 20. G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky. «Detection algorithm and initial laboratory results using V-BLAST space–time communication architecture,» *Electron. Lett.*, vol. 35, no. 1, pp. 14–16, Jan. 1999.
 21. R. L. Cupo, G. D. Golden, C. C. Martin, K. L., Sherman, N. R. Sollenberger, J. H. Winters, P. W. Wolniansky. «A Four-Element Adaptive Antenna Array For IS-136 PCS Base Stations», 1997 IEEE 47th Vehicular Technology Conference, May 1997.

УДК 621.391.26

МОДИФИКАЦИЯ БАНКА ФИЛЬТРОВ ВЕЙВЛЕТНЫХ ПРЕОБРАЗОВАНИЙ

Л.В. Новиков

Введение

Благодаря исследованиям в области математики, выполненным в конце 80-х годов, появился новый инструмент для обработки сигналов – вейвлеты. Они производят, по существу, частотно-локальный анализ и особенно эффективны для обработки финитных сигналов и обнаружения аномалий на фоне некоторого стационарного процесса. Для цифровой обработки сигналов чаще всего применяется дискретные вейвлетные преобразования (ДВП), построенные на основе теории кратномасштабного (мультиразрешающего) анализа (КМА) и реализуемые с помощью быстрых вычислительных алгоритмов [1–3]. Широкое применение ДВП нашло в различных приложениях, например, для аппроксимации и сжатия сигналов [4], обработки изображений [5], при измерении тока пучка ионов на тераваттном ускорителе [6], в системах регулирования быстропротекающими процессами, когда по результатам обработки принимаются необходимые управляющие решения [7], в аналитической химии [8], радиолокации [9] и др. областях [10,11].

КМА основан на идее использования масштабирующих $\varphi(t)$ и вейвлетных $\psi(t)$ функций, сдвиги и масштабирование которых образуют базисы для сигналов с соответствующим частотным спектром [1–3]. Масштабирующие функции

$$\left\{ \varphi_{j_0,k}(t) = 2^{\frac{j_0}{2}} \varphi(2^{j_0} t - k) \right\}$$

обеспечивают грубое представление сигнала в некотором пространстве низкочастотных колебаний (включая $\omega = 0$) при значении масштаба j_0 . Это представление затем улучшается дополнением высокочастотных деталей сигнала из вейвлетных пространств с базисом в каждом из них

$$\left\{ \psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k) \right\}$$

при значениях масштаба $j = j_0, j_0 + 1, \dots, J$.

Тогда любой сигнал с ограниченным частотным спектром может быть представлен в виде

Предложен подход расширяющий функции классического банка фильтров быстрых вейвлетных преобразований. Подход основан на использовании аппаратно – ориентированных вейвлетов приспособленных к решению определенного круга задач с учетом априорной информации об импульсном отклике измерительного тракта. Новый банк фильтров решает задачу оценки полезного сигнала (или его линейных преобразований) искаженного измерительным трактом и шумом. Применение предложенного подхода позволяет более чем на порядок повысить скорость обработки данных по сравнению с комбинированными алгоритмами

$$x(t) = \sum_{k=-\infty}^{\infty} c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_{j=j_0}^J \sum_{k=-\infty}^{\infty} d_j(k) \psi_{j,k}(t), \quad (1)$$

где первое слагаемое является грубым приближением сигнала, второе – его уточнениями вплоть до самого высокого разрешения при значении масштаба J ; $c_{j_0}(k)$ и $d_j(k)$ – коэффициенты разложения сигнала по масштабирующим и вейвлетным функциям соответственно; j_0 – значение масштаба грубого разрешения.

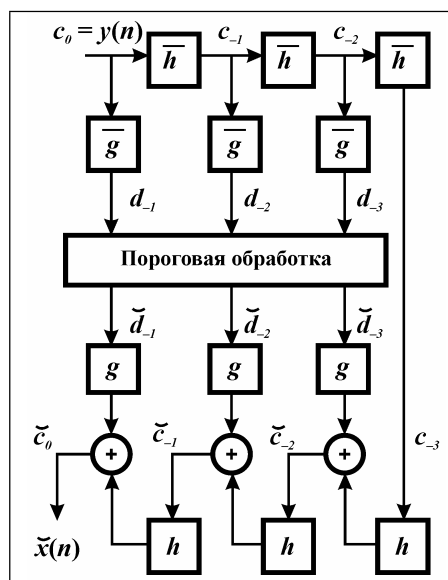
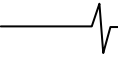


Рис1. Банк фильтров обработки сигналов, синтезированный на основе ортонормированных вейвлетов (например, вейвлетов Добеши).

Функция $\varphi(t)$ должна удовлетворять масштабирующему уравнению



$$\varphi(t) = \sum_n h(n) \sqrt{2} \varphi(2t - n), \quad (2a)$$

а вейвлетные функции – соотношению

$$\psi(t) = \sum_n g(n) \sqrt{2} \varphi(2t - n), \quad (2b)$$

где h и g – масштабирующие и вейвлетные коэффициенты (их еще называют, соответственно, низкочастотным и высокочастотным фильтрами), причем

$$g(n) = (-1)^{1-n} h(1-n) \quad (3)$$

Рассмотрим одну из распространенных задач, решаемых с помощью вейвлетной обработки. Пусть $y(t)$ – наблюдаемый сигнал, представляющий собой смесь полезного сигнала $x(t)$ и шума $u(t)$:

$$y(t) = x(t) + u(t)$$

Требуется оценить функцию $x(t)$. Решение этой задачи предложено в работе [12] и состоит из трех этапов: ДВП – пороговая обработка – обратное ДВП. Пример трехкаскадного банка фильтров, реализующего эту процедуру обработки, показан на рис.1. Вычисление вейвлетного спектра, содержащего низкочастотные

$$\{c_j(k)\}$$

и высокочастотные

$$\{d_j(k)\}$$

компоненты (где масштаб $j = 0, -1, -2, \dots$; и $k \in Z$ – положение спектральной составляющей) при каждом значении масштаба j осуществляется с помощью каскадного алгоритма Малла [2,3]:

$$c_j(k) = \sum_n h(n-2k) c_{j+1}(n) = \quad (4)$$

$$= c_{j+1} * \bar{h}(2k)$$

$$d_j(k) = \sum_n g(n-2k) c_{j+1}(n) = \quad (5)$$

$$= c_{j+1} * \bar{g}(2k)$$

где $(*)$ обозначает свертку;

$$\bar{h}(k) = h(-k), \bar{g}(k) = g(-k).$$

После выполнения операции свертки осуществляется прореживание отсчетов вдвое (децимация).

Этот алгоритм осуществляет преобразование «от тонкого к грубому разрешению» [3] и при самом тонком значении масштаба $j = 0$ полагают коэффициенты $c_0(n) = y(n)$, т.е. они являются отсчетами наблюдаемого сигнала.

В вейвлетной области при удачном выборе вейвлетного базиса подчеркиваются полезные компоненты наблюдаемого сигнала. Поэтому целью пороговой обработки является удаление компонент сигнала очень низкого уровня, не влияющие на конечный результат и компонент шума, энергия которого обычно равномерно распределяется по всему спектру. В результате происходит сжатие сигнала: число оставшихся спектральных составляющих значительно меньше объема массива ис-

ходных данных $y(n)$.

Восстановление сигнала (см. рис.1) производится по рекуррентной формуле

$$\begin{aligned} \tilde{c}_{j+1}(k) &= \\ &= \sum_n \tilde{c}_j(n) h(k-2n) + \sum_n \tilde{d}_j(n) g(k-2n) = \end{aligned} \quad (6)$$

$$= \tilde{c} * h(k) + \tilde{d} * g(k),$$

где $\tilde{c}, \tilde{d}$ – низкочастотные и высокочастотные спектральные компоненты после пороговой обработки;

$$\tilde{c}(n) = \begin{cases} \tilde{c}(k), & n = 2k \\ 0, & n = 2k + 1 \end{cases},$$

$$\tilde{d}(n) = \begin{cases} \tilde{d}(k), & n = 2k \\ 0, & n = 2k + 1 \end{cases},$$

т.е. до выполнения операции свертки производится увеличение числа отсчетов вдвое прореживанием коэффициентов $\tilde{c}(k)$ и $\tilde{d}(k)$ нулями (интерполяция).

При $j = -1$ получаем оценку полезного сигнала $x(t)$ в виде отсчетов $\tilde{x}(n) = \tilde{c}_0(n)$.

Однако в практике обработки сигналов часто требуется решение более сложной задачи: оценка сигнала $x(t)$ или его линейного преобразования при наличии двух видов искажений, вызванных шумом и измерительной системой с известным импульсным откликом (приборной функцией) $H(t)$:

$$y(t) = \int_{-\infty}^{\infty} x(\tau) H(t-\tau) d\tau + u(t). \quad (7)$$

Решение этой задачи обычно осуществляется комбинированными алгоритмами, включающими ДВП в качестве составной части, и фильтр Винера, дифференцирующий или другой фильтр в зависимости от того, какое преобразование над сигналом необходимо осуществить [см., например, 13]. Вместо этих алгоритмов в настоящей работе рассматривается подход, основанный на применении вейвлетов, синтезированных под решаемую задачу, т.н. аппаратно – ориентированных вейвлетов [14]. Банк фильтров, построенный на основе этих вейвлетов, обеспечивает одновременно получение оценки линейного преобразования сигнала, подавление шума и влияния приборной функции. В зависимости от постановки задачи обработки может быть выбран один из двух способов синтеза таких банков фильтров, рассматриваемых ниже.

Банк фильтров модифицированных вейвлетов

Приведем сначала некоторые расчетные соотношения. Предположим, что оценка линейного преобразования сигнала в модели (7) осуществляется с помощью фильтра $G(t)$. Тогда

$$\tilde{x}(t) = (G * y)(t), \quad (8)$$

где $G(t)$ выбирается в зависимости от решаемой задачи: в частности, это может быть фильтр Винера – Тихо-

нова, определяемый через преобразование Фурье как

$$\widehat{G}(\omega) = \frac{\widehat{H}(\omega)}{|\widehat{H}(\omega)|^2 + \beta},$$

где β – параметр регуляризации.

Решение ищем в виде

$$\begin{aligned} \tilde{x}(t) = & \sum_{k=-\infty}^{\infty} \tilde{c}_{j_0}(k) \varphi_{j_0,k}(t) + \\ & + \sum_{j=j_0}^J \sum_{k=-\infty}^{\infty} \tilde{d}_j(k) \psi_{j,k}(t) \end{aligned} \quad (9)$$

$$\text{где } \varphi_{j_0,k}(t) = 2^{\frac{j_0}{2}} \varphi(2^{j_0} t - k),$$

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k) -$$

известные ортонормированные масштабирующие и вейвлетные функции (например, вейвлеты Добеши [1]).

Для того чтобы найти оценку $\tilde{x}(t)$ по формуле (9), необходимо определить коэффициенты $\tilde{c}_{j_0}(k)$ и $\tilde{d}_j(k)$ по наблюдаемому сигналу $y(t)$.

Из формулы (9) с учетом (8) наблюдаемый сигнал можно представить в виде

$$\begin{aligned} y(t) = & \sum_{k=-\infty}^{\infty} \tilde{c}_{j_0}(k) \theta_{j_0,k}(t) + \\ & + \sum_{j=j_0}^J \sum_{k=-\infty}^{\infty} \tilde{d}_j(k) \eta_{j,k}(t) \end{aligned} \quad (10)$$

$$\text{где } \theta_{j,k}(t) = 2^{\frac{j}{2}} \theta_j(t - 2^{-j} k) \text{ и}$$

$$\eta_{j,k}(t) = 2^{\frac{j}{2}} \eta_j(t - 2^{-j} k) \quad (j, k \in \mathbb{Z})$$

масштабирующие и вейвлетные функции, Фурье – образы которых определяются выражениями

$$\widehat{\theta}_j(\omega) = \widehat{G}^{-1}(\omega) \widehat{\varphi}(2^{-j} \omega), \quad (11a)$$

$$\widehat{\eta}_j(\omega) = \widehat{G}^{-1}(\omega) \widehat{\psi}(2^{-j} \omega). \quad (11b)$$

Существуют двойственные базисные функции такие, что

$$\widehat{\widetilde{\theta}}_j(\omega) = \widehat{G}(\omega) \widehat{\varphi}(2^{-j} \omega), \quad (11b)$$

$$\widehat{\widetilde{\eta}}_j(\omega) = \widehat{G}(\omega) \widehat{\psi}(2^{-j} \omega). \quad (11g)$$

Функции (11) образуют семейство модифицированных вейвлетов, т.к. получены модификацией известных вейвлетов импульсным откликом фильтра $G(t)$.

Функции $\theta_{j,k}(t)$ и $\widetilde{\theta}_{j,k}(t)$ биортогональны по отношению к собственным сдвигам, функции $\eta_{j,k}(t)$ и $\widetilde{\eta}_{j,k}(t)$ биортогональны по отношению к сдвигам и масштабу, а все они обеспечивают кратномасштабный анализ сигналов из $L^2(\mathbb{R})$. Биортогональными являются также функции $\theta_{j,k}(t)$ и $\widetilde{\eta}_{j,k}(t)$, $\theta_{j,k}(t)$ и $\eta_{j,k}(t)$.

Учитывая это обстоятельство, можно показать, что функции $\widetilde{\theta}_{j,k}(t)$ и $\widetilde{\eta}_{j,k}(t)$ должны удовлетворять соотношениям, аналогичным (2), с масштабирующими и вейвлетными коэффициентами, которые обозначим соответственно, как $\widetilde{\alpha}_i(n)$ и $\widetilde{\beta}(n)$. В работе [15] показано, что в этом случае спектральные компоненты $\widetilde{c}_j(k)$ и $\widetilde{d}_j(k)$ вычисляются по рекуррентным формулам (4) и (5) кроме первого каскада, где они определяются как:

$$\begin{aligned} \widetilde{c}_{-1}(k) &= \sum \widetilde{\alpha}_0(m - 2k) y(m) \\ \widetilde{d}_{-1}(k) &= \sum_m \widetilde{\beta}_0(m - 2k) y(m) \end{aligned}$$

Синтез оценки полезного сигнала так же как и в случае обычного вейвлет-анализа производится по рекуррентной формуле (6).

Банк фильтров, реализующий процедуру вейвлетной обработки по оценке полезного сигнала $\tilde{x}(t)$ в модели (8) показан на рис.2. Он отличается от банка фильтров рис.1 только коэффициентами фильтров в первом каскаде.

Коэффициенты $\widetilde{\alpha}_0$ и $\widetilde{\beta}_0$ вычисляются по формулам

$$\widetilde{\alpha}_0(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{G(\omega)} \widehat{h}(\omega) \exp(i\omega n) d\omega, \quad (12a)$$

$$\widetilde{\beta}_0(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{G(\omega)} \widehat{g}(\omega) \exp(i\omega n) d\omega, \quad (12b)$$

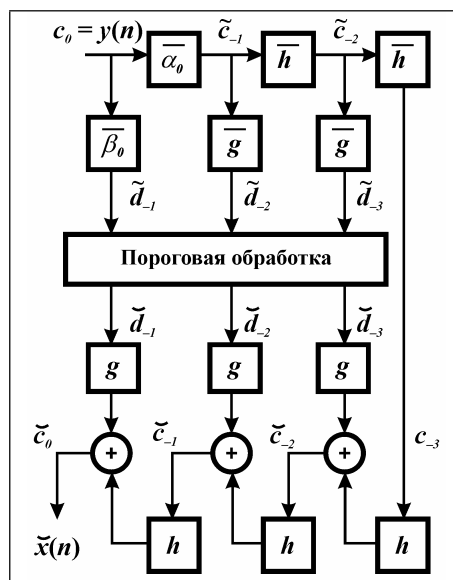


Рис. 2. Банк фильтров обработки сигналов, синтезированный на основе ортонормированных вейвлетов, модифицированных импульсным откликом линейной системы.

Если же функция $\widehat{G}(\omega)$ представляет собой частотный отклик дискретного фильтра, тогда справедливо равенство

$$\widehat{G}(\omega) = \widehat{G}(\exp(i\omega)) = G_0(z),$$

где $z = \exp(i\omega)$. Так как $h(n)$ и $g(n)$ также являются дискретными фильтрами, то коэффициенты $\widetilde{\alpha}_0(n)$ и

$\tilde{\beta}_0(n)$ могут быть получены путем перемножения их Z - преобразований:

$$\tilde{\alpha}_0(z) = (\tilde{G} * h)(z) \quad (12в)$$

$$\tilde{\beta}_0(z) = (\tilde{G} * g)(z) \quad (12г)$$

Тогда искомые коэффициенты будут коэффициентами z - преобразований $\tilde{\alpha}_0(z)$ и $\tilde{\beta}_0(z)$.

Если необходимо вычислить, например, p -ю производную оценки полезного сигнала (с подавлением влияния аппаратной функции), то коэффициенты $\tilde{\alpha}_0(z)$ и $\tilde{\beta}_0(z)$ необходимо определять по формулам (12а) и (12б), помножив функцию $G(\omega)$ на $(i\omega)^p$. Если же необходимо вычислить производную наблюдаемого сигнала, то коэффициенты $\tilde{\alpha}_0(z)$ и $\tilde{\beta}_0(z)$ необходимо определять по формулам (12в) и (12г), положив в этих формулах в качестве функции $\tilde{G}$ частотный отклик некоторого дифференцирующего дискретного фильтра.

Банк фильтров квазивейвлетов

С целью решения задачи оценки функции $x(t)$ в модели (7) может быть применен другой подход, основанный на прямом синтезе вейвлетов с использованием импульсного отклика $H(t)$.

Как отмечено в [1], кратномасштабный анализ можно построить с использованием некоторых функций $f(t)$, обладающих свойством частотно-временной локализации, сдвиги которых с относительным шагом b_0 формируют базис Рисса $\{f_k(t) = f(t - b_0 k), k \in Z\}$ и удовлетворяющих масштабирующему уравнению (2а). Ограничения, налагаемые на «кандидатов» в масштабирующие функции обсуждаются в работе [16]. Однако функции $f(t)$, особенно выбранные из некоторых практических соображений, удовлетворяют требованиям, предъявляемым к масштабирующим функциям, с некоторым приближением. Очевидно, что для таких функций равенство в (2) также будет приближенным. Именно поэтому масштабирующие и вейвлетные функции, синтезированные в этих условиях целесообразно назвать квазивейвлетами. Как и ранее приведем некоторые расчетные соотношения для синтеза коэффициентов банка фильтров.

Рассматриваемый подход применен в том случае, когда импульсный отклик линейной системы (прибора, экспериментальной установки и т.п.) $H(t)$ можно аппроксимировать функцией, заданной аналитически и имеющей экспоненциально убывающие фронты (т.е. имеющей вид пика). Тогда, положив

$$f(t) = \frac{H(-t)}{\int_{-\infty}^{\infty} |H(t)|^2 dt},$$

путем сдвига с шагом b образуем множество функций

$$\left\{ f_k(t) = f\left(\frac{t - bk}{\mu}\right) = f(t - b_0 k) \right\}, k \in Z,$$

где $b_0 = b/\mu$, μ – среднеквадратическая ширина или

ширина на половине высоты (полуширина).

Выберем относительный шаг b_0 из интервала $0,8 \dots 1,6$ и воспользуемся вавшись модифицированным способом ортогонализации Грамма–Шмидта, предложенным в [17], образуем из системы функций $\{f_k(t)\}$ ортонормированный базис $\{\varphi_k(t) = \varphi(t - b_0 k)\}$:

$$\varphi(t) = \sum_m \gamma(m) f(t - b_0 m), \quad (13)$$

где γ – коэффициенты ортогонализации.

Одним из инженерных способов проверки «удачного» выбора b_0 является быстрое убывание γ

Коэффициенты масштабирующего уравнения h вычисляем по формуле

$$h(n) = \int_{-\infty}^{\infty} \varphi(t) \varphi(2t - b_0 n) dt$$

Коэффициенты $g(n)$ вычисляются по формуле (3).

Дополнительно можно проверить удовлетворяют ли коэффициенты $h(n)$ и $g(n)$ некоторым обязательным соотношениям [3]:

ортгональность по отношению к двойному сдвигу

$$\sum_n h(n) h(n - 2k) = \delta(k) = \begin{cases} 1, & \text{если } k = 0 \\ 0, & \text{если } k \neq 0; \end{cases}$$

нормировка

$$\sum_n h(n) = 1;$$

Можно допустить, что все приведенные выше равенства должны соблюдаться с точностью, например, до четвертого знака после запятой. Коэффициенты γ , h и g зависят только от выбранной функции $H(t)$, описывающей импульсный отклик, и параметра сдвига b_0 и не изменяются при сохранении настройки прибора.

Покажем теперь, что синтезированный выше базис решает задачу оценки $x(t)$. Пусть сигнал $x(t)$ и масштабирующие функции $\{\varphi_k(t), k \in Z\}$ принадлежат одному подпространству. Тогда справедливо равенство

$$x(t) = \sum_k c_0(k) \varphi_k(t). \quad (14)$$

С учетом (13) для коэффициентов $c_0(k)$ получим вычислительный алгоритм

$$\begin{aligned} c_0(k) &= \int_{-\infty}^{\infty} x(t) \varphi_k(t) dt = \\ &= \sum_n y(bn) \gamma(n - k) = \sum_n y(bn) \bar{\gamma}(k - n) \end{aligned}$$

где $\bar{\gamma}(k) = \gamma(-k)$;

$$y(bn) = y\left(\frac{b}{\mu} n\right) = \frac{1}{\int_{-\infty}^{\infty} |H(t)|^2 dt} \int_{-\infty}^{\infty} x(t) H\left(\frac{bn - t}{\mu}\right) dt$$

– отсчеты выходного сигнала прибора, взятые с интервалом $b = b_0 \mu$, $k, n \in N$. Так как интервал дискретизации наблюдаемого сигнала из-за наличия шума должен быть меньше b , вычисление $c_0(k)$ должно быть выпол-

нено для всех временных отсчетов $y(n)$.

Чтобы получить алгоритм вычисления искомой оценки $\tilde{x}(t)$ по коэффициентам $\tilde{c}_0(k)$, из формулы (14) с учетом (13) получим:

$$\begin{aligned}\tilde{x}(t) &= \sum_k \tilde{c}_0(k) \varphi_k(t) = \\ &= \sum_n f\left(\frac{t - bn}{\mu}\right) \sum_k \tilde{c}_0(k) \gamma(n - k).\end{aligned}$$

Следовательно, для полного восстановления сигнала $x(t)$ необходимо $\tilde{c}_0(k)$ пропустить через фильтр $\gamma(k)$, после чего, воспользовавшись функцией $f(t)$, получаем искомую оценку $\tilde{x}(t)$.

Банк фильтров реализующий рассматриваемый алгоритм, показан на рис. 3. Полезный сигнал $x(t)$, который требуется оценить, искаженный приборной функцией $H(t)$ и шумом в виде дискретных временных отсчетов $y(n)$ поступает на вход дискретного фильтра с откликом γ . Выход этого фильтра подвергается ДВП фильтрами h и g (на рис.3 показаны три каскада ДВП) с последующей пороговой обработкой, после чего выполняется обратное ДВП фильтрами h и g . Выходная дискретная последовательность обрабатывается фильтром γ и затем, с помощью фильтра с импульсным откликом $f(t)$ вычисляется искомая оценка сигнала $\tilde{x}(t)$. Коэффициенты фильтров γ , h и g находятся по формулам, приведенным выше.

В практике обработки сигналов, как правило, возникает необходимость регуляризации решения задачи деконволюции с целью подавления ложных выбросов и автоколебаний. Этого можно достичь путем корректировки усиления верхних частот (аналогичный параметр имеется в фильтре Винера-Тихонова [13]) и частотной настройки банка фильтров.

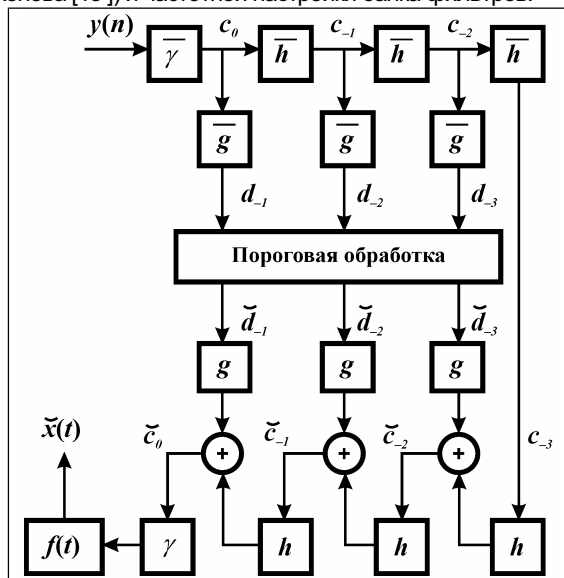


Рис. 3. Банк фильтров обработки сигналов на основе квазивейвлетов

Коррекция усиления верхних частот производится следующим образом:

- построить фильтр нижних частот $w(m)$ с коэффициентами, полученными, например, как
- $$w(m) = \exp\left\{-m^2/2\beta^2\right\}$$

для $m = 0, 1, 2, \dots$, где β – параметр регуляризации;

- нормировать коэффициенты $w_0(m) = w(m) / \sum w(m)$;
- выполнить свертку коэффициентов γ с $w_0(m)$;

Частотная настройка производится путем варьирования шагом отбора b отсчетов $y(bn)$ на входе, и, соответственно, шириной функции $f(t)$ (так как $b = b_0 \mu$) на выходе банка фильтров рис.3.

Приведенные выше подходы для синтеза банка фильтров является одним из способов адаптации вейвлетов к решению конкретных задач обработки сигналов. Их применение позволяет более чем на порядок повысить скорость обработки данных по сравнению с комбинированными алгоритмами.

Список литературы

1. Daubechies I. Ten lectures on Wavelets. Philadelphia, PA: SIAM, 1992 (Перевод: Добеши И. Десять лекций по вейвлетам. М.: РХД, 2002.)
2. Mallat S. A wavelet Tour of Signal Processing. San Diego: Academic Press, 1999. (Перевод: Малла С. Вейвлеты в обработке сигналов. М.: Мир, 2005.)
3. Burrus C.S., Gopinath R.A., Haitao Guo. Introduction to Wavelets and Wavelet Transform. New Jersey: Prentice Hall, 1998.
4. Vetterly M. Wavelets, Approximation, and Compression // IEEE Signal Proc. Magazine, 2001. September. P.59.
5. Shen L., Sun Q. Biorthogonal Wavelet System for High-Resolution Image Reconstruction // IEEE Transaction on Signal Processing. 2004. V.52, №7. P.1997.
6. Борисенко Н.А., Фертман А.Д. Автоматизированный анализ экспериментальных данных с применением вейвлет-преобразования // ПТЭ, 2003. №2. С.28.
7. Шищенко В.А., Любимов В.В., Иванова Т.И. Автоматизация и современные технологии, 2002. №1. С.3.
8. Shao X.-G., Leung A. K.-M., Chau F.-T. Wavelet: A New Trend in Chemistry // Accounts of Chemical Research. 2003. V.36, №4. P. 276-283.
9. Бессонов Д.П., Сизов А.С. Проблемы автоматического распознавания сложных радиотехнических сигналов и пути их решения на основе применения вейвлет-преобразования // Изв. вузов. Приборостроение. 2005. №2. С.31–34.
10. Дремин И. М., Иванов О.В., Нечитайло В.А. Вейвлеты и их использование. // УФН. 2001. Т.171, №5. С.465.
11. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения. // УФН. 1998. Т.166, №11. С.1145.
12. Donoho D., Johnstone I. Ideal spatial adaptation via wavelet shrinkage. // Biometrika. 1994. V.81. P. 425.
13. Neelamani R., Choi H., Baraniuk R.G. Wavelet – Domain Regularized Deconvolution for Ill – Conditioned Systems. // IEEE Transaction on Signal Processing, 2004. V.52, №2. P.418.
14. Новиков Л.В. Аппаратно – ориентированные вейвлеты и их применение в обработке экспериментальных данных// ПТЭ. 2005. №6. С.13
15. Новиков Л.В. Модифицированные вейвлеты в обработке данных аналитических приборов. I. Основы теории // Научное приборостроение. 2006. Т. 16. №1. С.
16. Strang G., Zhou D.-X. The Limits of Refinable Functions// Trans. American Math. Soc. 2001. V. 353. P.1971-1984
17. Новиков Л.В. Адаптивный вейвлет-анализ сигналов // Научное приборостроение. 1999. Т. 9. №2. С.30.

НЕЙРОСЕТЕВОЙ МЕТОД ЦИФРОВОЙ ОБРАБОТКИ НЕСТАЦИОНАРНОГО СИГНАЛА (РЕЧИ) В ВЕЙВЛЕТ-ОБЛАСТИ

А.В. Меркушева

Основные принципы

Введение. Детектирование момента изменения свойств нестационарного сигнала является необходимым условием фильтрации с адаптивной подстройкой порога, величина которого зависит от уровня шума. Такая воз-

можность предоставляется в информационно-измерительной системе (ИИС), где зашумленный сигнал, несущий полезную информацию, регистрируется на временных интервалах, которые перемежаются с интервалами, свободными от сигнала. Типичный случай этого варианта контроля и обработки реализуется в информационной системе анализа речевого сигнала. Система служит моделью достаточной общности для нестационарных сигналов в ИИС со спорадически появляющимися информационно значимыми интервалами различной длительности. Другим примером системы такого вида является ИИС, предназначенная для контроля гетерогенного потока товарной нефти [1], где выполняется выделение временных сегментов, на которых отсутствует свободный газ, определяется средняя плотность потока при наличии и отсутствии газа, а затем на основе этих данных вычисляется средняя относительная доля свободного газа.

Метод детектирования изменения свойств нестационарного сигнала основан на вейвлет-преобразовании и нейросетевом алгоритме и рассматривается применительно к информационной системе многоцелевой обработки речевого сигнала (РС). Адаптивная фильтрация шума реализуется путем динамического изменения порога, вычисляемого на интервалах (микропаузах) отсутствия РС, поэтому правильное определение наличия или отсутствия речевого сигнала на анализируемом интервале оказывает существенное влияние на качество фильтрации. Эта задача является актуальной в различных системах обработки РС:

- в цифровой телефонии для удаления пауз при сжатии речевых сигналов;
- в криптографии, где удаление пауз сокращает избыточность, уменьшающую криптостойкость алгоритма шифрования;
- в устройствах фильтрации речевого сигнала с адаптацией к виду и уровню шума, где во время пауз должны быть получены кратковременные характеристики шума.

Таким образом, в описанных ИИС и других приложениях необходимо детектирование (отслеживание и идентификация) изменения свойств нестационарных процессов, а также адаптация к изменению уровня шума. Это особенно важно для обработки сигнала в реальном времени, при

Рассмотрены основные принципы нейросетевого метода детектирования изменения свойств нестационарного речевого сигнала (РС) в области вейвлет-разложения. Представлены результаты разработки и исследования нейросетевого метода цифровой обработки РС: выбор сложности многослойного персептрона на нейронной сети (НС); определение лучшего алгоритма нелинейной оптимизации для обучения НС; способ снижения размерности входных векторов при обучении НС, основанный на принципе главных компонент.

которой алгоритм обработки должен адаптироваться к состоянию процесса.

В разработанном методе использовано вейвлет-разложение речевого сигнала по биортогональному базису Добеши [2], а для принятия решения о типе интервала (сегмента) сигнала – нейронная сеть на многослойном персептроне. Сравнительно с традиционными спектральными методами вейвлет-преобразование дает более точную локализацию сигнала по времени и по частоте (в субполосах разложения), имеет быстрый алгоритм реализации. Биортогональный базис сохраняет фазовые соотношения частотных компонент сигнала после его восстановления обратным вейвлет-преобразованием. Свойство наличия у базиса преобразования конечного числа нулевых моментов гарантирует точную аппроксимацию полиномиальной части сигнала, что ведет к дополнительному снижению объема данных [3].

Вейвлет-разложение речевого сигнала в соответствии с персептуальной моделью. С целью адаптации цифровой обработки сигнала к переменному во времени шуму, при решении задачи распознавания речи и пауз предложен подход на основе вейвлет-преобразования и нейронной сети [7, 8]¹. Вейвлет-преобразование позволяет более точно локализовать частотные свойства сигнала во времени. Биортогональные вейвлеты Добеши позволяют уменьшить объем вычислений при разложении за счет использования коротких фильтров. Вейвлет-функция и масштабированная (скейлинг) функция, используемые для разложения сигнала, представлены на рис.1. Вейвлет-разложение не приводит к увеличению объема данных при переходе от временного представления сигнала к его представлению в вейвлет-области [9].

Алгоритм разделения речи и пауз построен с учетом особенностей восприятия звука человеком [10,11,12], которые описываются персептуальной моделью [13,14,15]. Модель разделяет спектр речевого сигнала на частотные полосы, называемые критическими. Каждая полоса частотного диапазона речи по компоненте шума в ней воспринимается как единое целое, и по слуховому ощущению важна лишь мощность шума в полосе [16]. Персептуальная модель успешно используется в алгоритмах MPEG (Moving

Picture Expert Group по стандартизации аудио- и видеоин-

формации) для кодирования акустических сигналов.

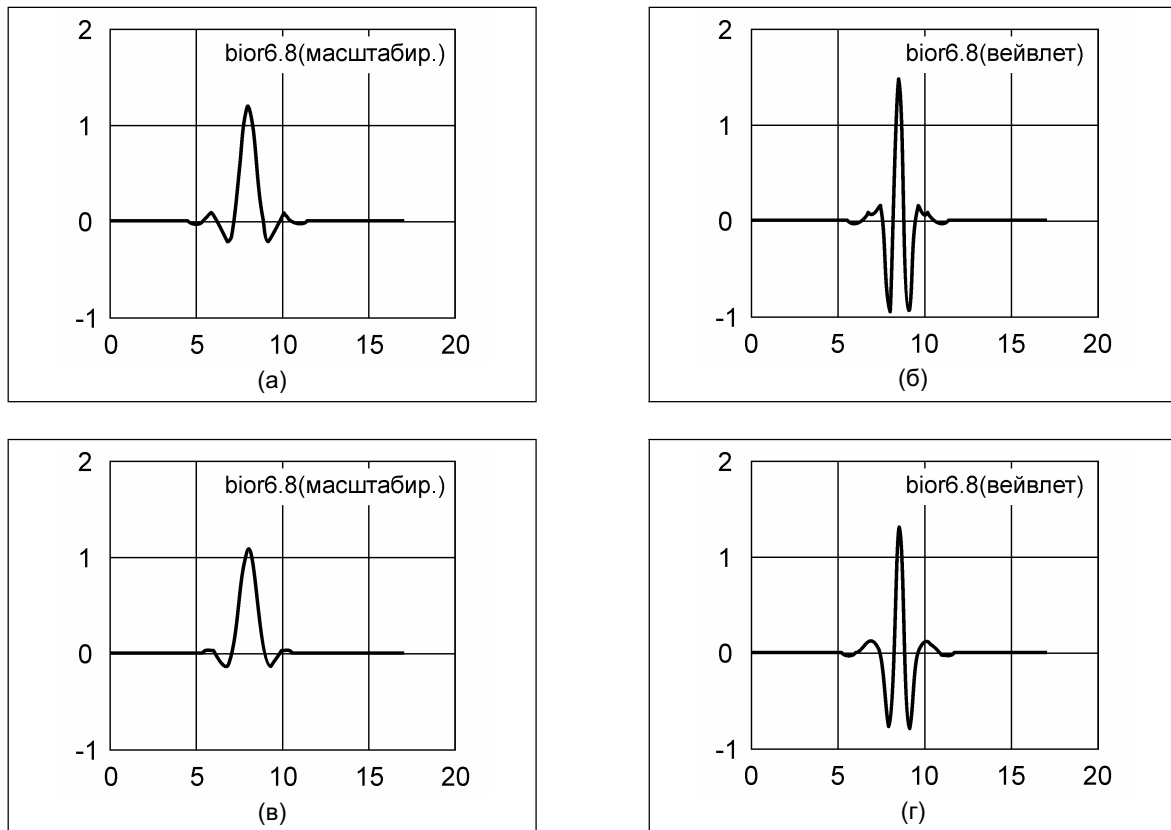


Рис.1. Базисные функции, используемые при разложении сигнала с помощью вейвлет-пакетов: а – масштабирующая функция анализа; б – вейвлет-функция анализа; в – масштабирующая функция синтеза; г – вейвлет-функция синтеза

По перцептуальной модели частотному диапазону 100–12000 Гц соответствует 22 критические полосы, ширина которых постепенно возрастает от 100 до 2500 Гц (табл.1). Диапазон РС ограничен частотой 8 кГц, однако в компьютерных системах обработки РС для уменьшения шума, связанного с дискретизацией и квантованием, используют повышенную частоту дискретизации, в частности 22 кГц. Даже при 10%-й неточности оценки частотного диапазона речи, начиная от самых низких частот, – в пределах 50–9500 Гц, его полностью перекрывают 22 полосы перцептуальной модели. Поэтому в системе обработки РС следует использовать 22 полосы перцептуальной модели.

Для определения речевой активности предложен метод [17], который включает:

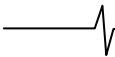
- получение сегмента речевого сигнала длительностью 10 – 20 мс;
- использование вейвлет-пакетов для получения разло-

жения сегмента в соответствии с перцептуальной моделью;

- вычисление мощности вейвлет-коэффициентов в каждой области разложения;
- нормирование вектора мощности вейвлет-коэффициентов, полученного на сегменте речевого сигнала;
- использование нормированного вектора в качестве входного вектора нейронной сети.

Вейвлет-преобразование сигнала производит октавополосное разбиение спектра, которое подходит для большинства (но не для всех) задач обработки сигнала. В частности, для получения перцептуальной модели необходимо изменить структуру разбиения частотно-временной плоскости. Достичь гибкого разбиения можно с помощью каскадного соединения блоков вейвлет-фильтров, которое реализуется в методе вейвлет-пакетов. Метод вейвлет-пакетов позволяет на каждом уровне разложения РС производить

¹ Изменение свойств речевого сигнала на интервалах речь/пауза обычно определяется изменением спектральных характеристик. В частности, алгоритм Рабинера (L. Rabiner) [4], разработанный для разделения временных интервалов речевой активности и пауз, построен на принципе анализа энергии сигнала и числа его переходов через нулевой уровень. Алгоритм не учитывает особенностей частотного спектра сигнала и шума и поэтому не всегда позволяет правильно классифицировать интервалы, особенно в условиях достаточно мощного узкополосного шума или музыкального фона. Применение спектрального анализа для детектирования свойств сигнала также не является эффективным, поскольку получение спектра Фурье связано с использованием всего интервала анализируемого сигнала, а кратковременный спектр Фурье имеет либо плохое разрешение по частоте, либо занимает большой временной интервал [5, 6]. Метод Фурье недостаточен также в связи с тем, что после фильтрации восстановление обратным преобразованием может привести к комплексной форме сигнала.



наиболее выгодное разбиение частотного диапазона [18–23]. Полученное разложение РС в соответствии с персептуальной моделью обеспечено специальным выбором двоичного дерева, определяющего структуру вейвлет–преобразования.

Таблица 1

Частотные диапазоны персептуальной модели

Индекс диапазона	Граничные частоты (Гц)	Средняя частота (Гц)	Ширина диапазона (Гц)
1	0–100	50	100
2	100–200	150	100
3	200–300	250	100
4	300–400	350	100
5	400–510	455	110
6	510–630	570	120
7	630–770	700	140
8	770–920	845	150
9	920–1080	1000	160
10	1080–1270	1175	190
11	1270–1480	1355	210
12	1480–1720	1600	240
13	1720–2000	1860	280
14	2000–2320	2160	320
15	2320–2700	2510	380
16	2700–3150	2925	450
17	3150–3700	3425	550
18	3700–4400	4050	700
19	4400–5300	4850	900
20	5300–6400	5850	1100
21	6400–7700	7050	1300
22	7700–9500	8750	1800
23	9500–12000	10750	2500
24	12000–15500	13750	3500
25	15500–23500	1950	8000

Критерием качества аппроксимации персептуальной модели служит показатель среднеквадратичной относительной погрешности соответствия границ критических полос и границ субполос вейвлет–пакета [7]. Этот показатель (Q) представлен соотношением (1):

$$Q = \sqrt{\frac{1}{p} \sum_{i=1}^p \frac{1}{2} \left\{ \left(\frac{\hat{f}_i^1 - \bar{f}_i}{\bar{f}_i} \right)^2 + \left(\frac{\hat{f}_i^2 - \bar{f}_i}{\bar{f}_i} \right)^2 \right\}}, \quad (1)$$

где $\hat{f}_i^1, \hat{f}_i^2$ – нижняя и верхняя границы частотной полосы в персептуальной модели,

$\bar{f}_i$ – положение центра полосы;

$\hat{f}_i^1, \hat{f}_i^2$ – то же для субполос сформированного вейвлет–пакета;

p – количество частотных диапазонов модели, p = 22. Несмотря на то что бинарная структура формирования дерева дает ограниченное количество вариантов, полученное разложение удовлетворительно аппроксимирует частотные диапазоны персептуальной модели. При первоначальной

оценке показателя Q=0.08, небольшое укрупнение трех полос в частотных диапазонах персептуальной модели (табл.2) улучшил общее соответствие представления речевого сигнала вейвлет–пакетом и при этом показатель качества Q снижен до 0.03.

Таблица 2

Сопоставление частотных субполос вейвлет–представления РС с персептуальной моделью

Верхние границы субполос и индексы частотных диапазонов вейвлет–пакета и персептуальной модели (Гц)			
Вейвлет–представление пакетом		Персептуальная модель	
Верхние границы субполос вейвлет–пакета	Индексы вершин дерева вейвлет–пакета ¹⁾	Верхние границы полос модели	Индексы полос ²⁾
250	1	300	1–3
375	2	400	4
500	3	510	5
625	4	630	6
750	5	770	7
1000	6	1080	8–9
1250	7	1270	10
1750	8–9	1720	11–12
2000	10	2000	13
2250	11	2320	14
2750	12–13	2700	15
3000	14	3150	16
3500	15–16	3700	17
4500	17–18	4400	18
5500	19	5300	19
6500	20	6400	20
7500	21	7700	21
9750	22	9500	22
12025	23	12000	23

¹⁾ Индексы 8–9, 12–13, 15–16 и 17–18 указывают на объединение субполос пакета на вершинах дерева разложения сигнала. Верхняя граница субполосы в первом столбце соответствует второму индексу из указанных пар.

²⁾ Индексы 1–3, 8–9, 11–12 соответствуют сгруппированным полосам персептуальной модели. Верхние границы полос соответствуют вторым индексам из указанных пар.

Само вейвлет–преобразование не дает требуемого уменьшения размерности (полное число коэффициентов разложения равно числу отсчетов сигнала), поэтому для нейросетевого алгоритма использован вектор с компонентами субполосной мощности сигнала $P = (\sigma_1^2, \sigma_2^2 \dots \sigma_p^2)$, где σ_i^2 – сумма квадратов коэффициентов, принадлежащих i-й полосе разложения.

Таким образом, с использованием вейвлет–пакетов для каждого сегмента РС получены векторы мощности вейвлет–коэффициентов в субполосах, соответствующих персептуальной модели. Эти векторы отражают спектральные свойства сигнала приблизительно с теми же градациями в частотной области, что и слуховой анализатор человека.

Нормирование векторов ($p = P / \|P\|$)

позволяет устранить влияние общей мощности сигнала на возможные решения о его свойствах и тем самым устранить основной недостаток алгоритмов детектирования РС, ориентированных в значительной степени на силу сигнала.

Применение персептрона для обнаружения свойств нестационарных сигналов. Для принятия решения о наличии или отсутствии речевого сигнала на анализируемом сегменте предложено применение нейронной сети [7],[8]. Входным вектором сети является нормированный вектор мощности вейвлет-коэффициентов. Такое решение имеет следующие преимущества:

- система, использующая нейронную сеть, является самонастраивающейся адаптивной системой, для которой нет необходимости предварительно выполнять точный расчет параметров;
- система имеет этап обучения, позволяющий выполнить адаптацию параметров для настройки на конкретные условия сигнала и шума;
- в области вейвлет-разложения не обязательно иметь линейную функцию для разделения вейвлет-образов

сигнала и шума, иными словами свойство линейной разделимости образов не является обязательным;

- нелинейная активационная функция нейрона позволяет принимать нежесткие решения в каждом нейроне сети;
- структура нейронной сети может быть адаптирована к задаче разделения сегментов сигнала и шума.

Нейронной сетью достаточно простой структуры является персептрон [24–27]. В простейшем случае нейроперсептрон состоит из одного нейрона (рис.2), у которого суммирующий узел дает линейную комбинацию входов x_i , приложенных к синапсам w_i , и учитывает внешний порог θ . К выходу суммирующего узла v ,

$$v = \sum_{i=1}^p w_i x_i - \theta, \text{ применяется нелинейная функция}$$

$y = \varphi(v)$, называемая функцией активации.

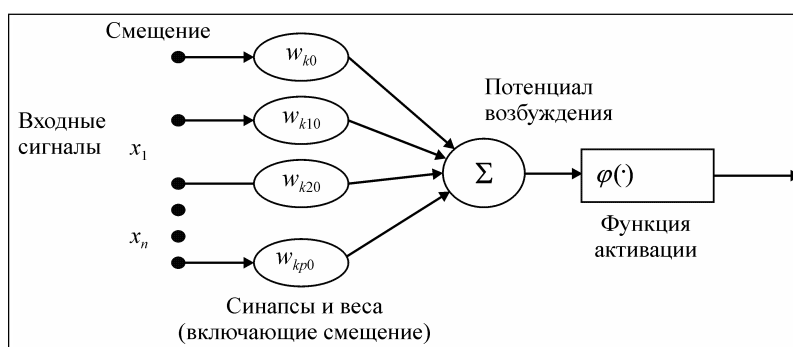


Рис.2. Структурная схема элементарного нейронного персептрона

Нейроперсептрон с одним слоем не имеет скрытых нейронов и не может классифицировать входные векторы, не являющиеся линейно разделимыми. Эта проблема решается введением скрытых слоев в многослойном персептроне (МСП), который состоит из входных узлов (составляющих входной слой), одного или более скрытых слоев и выходного слоя вычисляющих узлов. Входной сигнал распространяется через сеть слой за слоем. Каждый нейрон сети содержит нелинейную функцию активации в форме логистической кривой:

$$y_j = \varphi(v_j) = [1 + \exp(-v_j)]^{-1}.$$

Нейроперсептрон выполняет отображение входных p -мерных векторов

$P = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ на выходное q -мерное пространство, в задаче обнаружения РС $q = 1$.

Согласно Сайбенко–Фунахаши [28],[29], нейроперсептрон способен с любой точностью аппроксимировать желаемое отображение вход–выход, – обучаться сопоставлять множеству входных p -векторов заданное множество выходных q -векторов. Так что, существует ε -аппроксимация F любого непрерывного отображения вход–выход f , которую можно представить в математической форме (2), и эта форма эквивалентна двухслойному персептрону (3) с одним выходным нейроном и одним скрытым слоем из M нейронов (M зависит от величины ε):

$$|F(x_1, x_2, \dots, x_p) - f(x_1, x_2, \dots, x_p)| \leq \varepsilon, \quad (2)$$

$$F(x_1, x_2, \dots, x_p) = \sum_{j=0}^M a_j \varphi \left(\sum_{i=1}^p w_{ji} x_i - \theta_j \right) \quad (3)$$

Здесь a_j – коэффициенты аппроксимации отображения f обобщенным рядом Фурье, которые в персептроне представлены синаптическими связями (весами) выходного нейрона; $\{w_{ji}\}_{i,j=1..M}$ – синаптические веса нейронной сети от i -го к j -му нейрону, структурно связывающие входные узлы с нейронами скрытого слоя; θ_j – потенциал смещения (порог чувствительности) j -го нейрона; φ – функция преобразования потенциала возбуждения нейрона

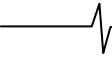
$$v_j = \left(\sum_{i=1}^p w_{ji} x_i - \theta_j \right) \text{ в его выходное значение } y_j.$$

Изложенное подтверждает правильность выбора типа нейронной сети в целях разработки метода для детектирования изменения свойств нестационарного сигнала.

Заключение. Предложенный метод и модель анализа речевого сигнала (РС) дают основу для решения ряда задач детектирования изменения свойств нестационарного сигнала:

1. Для анализа нестационарного сигнала (прототипом которого служил РС) целесообразно использование нейросетевых алгоритмов на вейвлет–отображении сигнала.

2. Вейвлет–разложение может быть адаптировано к частотной структуре сигнала на основе вейвлет–пакета. Для РС выбор вида пакета определен в соответствии с персептуальной моделью восприятия звука, которая построена на базе обобщения биофизических исследований процессов фильтрации органа слуха человека.



3. Для реализации нейросетевого алгоритма необходимо решение ряда вопросов:

- Обоснование метода обучения, имеющего лучшую скорость сходимости для векторов аудио-сигнала и позволяющего получить меньшие по величине ошибки распознавания свойств сигнала.
 - Получение ограничения на требуемое время обучения персептрона.
 - Обоснование оптимальной структуры персептрона, которая не является избыточной и позволяет эффективно решать задачи распознавания.
- Эти вопросы отражены во втором разделе этой статьи.

Исследование и оптимизация нейронной сети

Характеристики методов обучения нейроперсептрона. Для алгоритма обучения нейроперсептрона задаче детектирования речевой активности важно иметь быструю сходимость и малые вероятности ошибочных решений на этапе распознавания. Для этой цели применены различные алгоритмы, используемые при обучении нейронных сетей, и исследованы их характеристики для задачи распознавания типа сегментов речь/пауза. Алгоритмы построены на принципе минимизации суммарной ошибки распознавания и основаны на методах нелинейной оптимизации. Проанализированы метод сопряженных градиентов; простое и обобщенное правила дельта-дельта; метод случайного поиска; метод Бройдена – Флетчера и метод параллельных касательных (метод партан).

При обучении нейросети с настройкой ее параметров, – синаптических весов (СВ), применяется «алгоритм с обратным распространением ошибки» (алгоритм ОРО) [29 – 35]. Алгоритм использует понятия:

- ошибки $e_j(n)$ на выходе нейрона j на итерации n :

$$e_j(n) = d_j(n) - y_j(n),$$

где d_j – желаемый отклик, y_j – выход нейрона j , являющегося выходным узлом сети;

- текущего значения суммы квадратов ошибок:

$$E(n) = 0.5 \sum_{j \in C} e_j^2(n),$$

где C – множество выходных нейронов сети.

Цель обучения состоит в подстройке СВ сети, обеспечивающей минимум средней ошибки:

$$E_{cp} = (N)^{-1} \sum_{n=1}^N E(n),$$

где N – количество обучающих векторов.

Алгоритм ОРО реализует обучение итеративно с двумя проходами по сети – прямым и обратным, при которых соответственно вычисляются функциональные преобразования вектора входного сигнала $\mathbf{x} = (x_1, \dots, x_p)^T$ и корректировка вектора параметров $\mathbf{w}$. При обратном проходе на каждой итерации n определяется локальный градиент $\delta_j(n)$, для чего служит выражение (4), если нейрон находится в выходном слое, и выражение (5), если нейрон в скрытом слое:

$$\delta_j(n) = e_j(n) \phi'(v_j(n)) \quad (4)$$

$$\delta_j(n) = \phi'(v_j(n)) \sum_k \delta_k(n) w_{jk}(n), \quad (5)$$

где $v_j(n)$ – возбуждение j -го нейрона, $y_j(n)$ – выход j -го ней-

рона, $e_j(n)$ – ошибка на выходе j -го нейрона,

$$\phi'(v_j(n)) = y_j(n)[1 - y_j(n)] -$$

производная функции нелинейности, w_{jk} – синаптические веса. Локальный градиент ошибки $\delta_j(n)$ используется для корректировки весов в соответствии с δ -правилом по (6) или в соответствии с модифицированным δ -правилом по выражению (7):

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad (6)$$

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n), \quad (7)$$

где α и η – параметры, определяющие скорость обучения.

Непосредственное применение метода обратного распространения при реализации алгоритма обучения предполагает задание свободных параметров α и η .

Более продуктивный подход к обучению основан на нелинейной минимизации функции ошибок, или оптимизации нелинейной функции отображения вход-выход [36–39].

Метод сопряженных градиентов. В методе сопряженных градиентов [40] коррекция вектора весов персептрона $\mathbf{w}(n)$ осуществляется не вдоль вектора градиента ошибки $\mathbf{g}(n)$, а вдоль вектора сопряженного градиента $\mathbf{p}(n)$, который обновляется на каждом шаге итерации:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n) \mathbf{p}(n); \quad (8)$$

$$\mathbf{p}(n+1) = -\mathbf{g}(n+1) + \beta(n) \mathbf{p}(n),$$

где $\mathbf{g}(n)$ – вектор градиента функции ошибки $E(\mathbf{w}_n)$, а параметр $\beta(n)$ определяется по (9) для метода Флетчера–Ривса и по (10) – для метода Полака – Рибера [37]:

$$\beta(n) = \frac{\mathbf{g}^T(n+1) \mathbf{g}(n+1)}{\mathbf{g}^T(n) \mathbf{g}(n)} \quad (9)$$

$$\beta(n) = \frac{\mathbf{g}^T(n+1) [\mathbf{g}(n+1) - \mathbf{g}(n)]}{\mathbf{g}^T(n) \mathbf{g}(n)} \quad (10)$$

Оценка параметра $\eta(n)$ включает линейный поиск, состоящий в определении η , для которого критерий $E_{cp}[\mathbf{w}(n) + \eta \mathbf{p}(n)]$ имеет минимальную величину при данных значениях $\mathbf{w}(n)$ и $\mathbf{p}(n)$:

$$\eta(n) = \arg \left(\min_{\eta} \{E_{cp}[\mathbf{w}(n) + \eta \mathbf{p}(n)]\} \right) \quad (11)$$

Таким образом, при непосредственном использовании алгоритма ОРО параметр скорости обучения η определяется на основе оптимизации критерия минимума функции ошибок.

Метод дельта – дельта и обобщенное правило дельта – дельта. В качестве более совершенного метода выбора $\eta(n)$ применена форма адаптации параметра скорости обучения на основе правил Джексона: правила дельта-дельта адаптации и обобщенного правила дельта-дельта [41]. Оба правила основаны на принципе быстрого спуска по поверхности функции ошибок, в качестве которой служит суммарная квадратичная ошибка нейроперсептрона $E(n)$ на текущем шаге процесса обучения (12), которая в этом методе рассматривается в пространстве $\{\eta_i(n)\}_{i=1,2,\dots}$ параметров скорости обучения для синаптических весов $w_{ji}(n)$ нейронной сети:

$$E(n) = 0.5 \sum_j e_j^2(n) = 0.5 \sum_j [d_j(n) - y_j(n)]^2 \quad (12)$$

Определение градиента

$$\frac{\partial E}{\partial \eta_i(n)}$$

основано на правиле (13) дифференцирования сложной функции и общих соотношениях (14), (15):

$$\frac{\partial E}{\partial \eta_i(n)} = \frac{\partial E(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial \eta_i(n)} \quad (13)$$

$$v_j(n) = \sum_i w_{ji}(n) y_i(n) \quad (14)$$

$$w_{ji}(n) = w_{ji}(n-1) - \eta_i(n) \frac{\partial E(n-1)}{\partial w_{ji}(n-1)} \quad (15)$$

Соотношение (14) выражает возбуждение j -нейрона выходами y_i нейронов предшествующего слоя, а (15) – правило градиента для изменения параметров персептрона $w_{ji}(n)$. Соотношения (14) и (15) позволяют выразить $v_j(n)$ в виде:

$$v_j(n) = \sum_i y_i(n) [w_{ji}(n-1) - \eta_i(n)] \quad (16)$$

Это приводит к определению компонент (17), (18) и (19), необходимых для получения формы выражения (13), пригодной для вычислительного алгоритма:

$$\frac{\partial v_j(n)}{\partial \eta_i(n)} = -y_i(n) \frac{\partial E(n-1)}{\partial w_{ji}(n-1)} \quad (17)$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \phi'_j(v_j(n)) \quad (18)$$

$$\frac{\partial E(n)}{\partial y_j(n)} = -[d_j(n) - y_j(n)] = -e_j(n) \quad (19)$$

Полученная таким образом окончательная форма выражения

$$\frac{\partial E}{\partial \eta_i(n)} \text{ имеет вид:}$$

$$\frac{\partial E(n)}{\partial \eta_i(n)} = -\phi'_j(v_j(n-1)) e_j(n) y_i(n) \left[-\frac{\partial E(n-1)}{\partial w_{ji}(n-1)} \right] \quad (20)$$

Если учесть, что для градиента функции ошибки относительно весов справедливо (21):

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -\phi'_j(v_j(n)) e_j(n) y_i(n), \quad (21)$$

то получается конечная форма градиента в пространстве параметров скорости обучения:

$$\frac{\partial E}{\partial \eta_i(n)} = -\frac{\partial E(n)}{\partial w_{ji}(n)} \cdot \frac{\partial E(n-1)}{\partial w_{ji}(n-1)} \quad (22)$$

Таким образом, значение параметра скорости обучения η , обеспечивающее наискорейший спуск к минимуму поверхности ошибок в пространстве параметров $\{\eta_i(n)\}$, адаптивно обновляется по алгоритму в форме соотношения:

$$\Delta \eta_i(n+1) = -\gamma \frac{\partial E(n)}{\partial \eta_i(n)} = \gamma \frac{\partial E(n)}{\partial w_{ji}(n)} \frac{\partial E(n-1)}{\partial w_{ji}(n-1)}, \quad (23)$$

где γ – константа, контролирующая шаг адаптации параметра скорости обучения η .

$$\text{Значения } \frac{\partial E(n)}{\partial w_{ji}(n)} \text{ и } \frac{\partial E(n-1)}{\partial w_{ji}(n-1)}$$

(компоненты градиента функции ошибок на шаге n и $(n-1)$ в итеративной процедуре обучения) служат основой следующих правил, сформулированных Джекобсом:

– Если компонента градиента

$$\frac{\partial E(n)}{\partial w_{ji}(n)}$$

имеет одинаковый знак на двух итерациях, то $\Delta \eta_i(n+1) > 0$, т.е. адаптация увеличивает скорость обучения для веса w_{ji} .

– Если у градиента меняется знак на соседних итерациях, $\Delta \eta_i(n+1) < 0$, то процедура адаптации уменьшает параметр скорости обучения для веса w_{ji} .

Приведенная форма алгоритма на основе правила дельта–дельта проста в вычислениях, но может дать неуверенную адаптацию при очень малых и больших значениях градиента. Лучшие результаты дает более сложная модификация алгоритма – обобщенное правило дельта–дельта [41], которое требует большего объема вычислений. Адаптация параметра $\eta_i(n)$ для синаптической связи w_{ji} определяется выражениями (24) – (26):

$$\Delta \eta_{ji}(n+1) = \begin{cases} k & \text{при } S_{ji}(n-1) D_{ji}(n) > 0, \\ -\beta \eta_{ji}(n) & \text{при } S_{ji}(n-1) D_{ji}(n) < 0, \\ 0 & \text{иначе,} \end{cases} \quad (24)$$

где $D_{ji}(n)$ и $S_{ji}(n)$ определяются по (25) и (26):

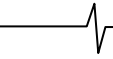
$$D_{ji}(n) = \frac{\partial E(n)}{\partial w_{ji}} \quad (25)$$

$$S_{ji}(n) = (1 - \xi) D_{ji}(n-1) + \xi \cdot S_{ji}(n-1) \quad (26)$$

В выражениях (25) и (26) ξ – положительная константа; $D_{ji}(n)$ – текущее значение градиента; $S_{ji}(n)$ – определяется соотношением типа авторегрессии, которое использует параметр веса ξ и компоненту градиента $D_{ji}(n-1)$ в качестве базы. Процедура адаптации параметра скорости обучения (24)–(26) составляет основу метода обучения по обобщенному правилу дельта–дельта. Адаптация основана на контроле совпадения или различия знаков на последовательных итерациях у соответствующих компонент $S_{ji}(n-1)$ и градиента $D_{ji}(n)$. При совпадении знаков $\Delta \eta_{ji}(n)$ составляет k , а при несовпадении уменьшается экспоненциально: $\Delta \eta_{ji}(n+1) = -\beta \eta_{ji}(n)$.

Метод случайного поиска. Метод случайного поиска основан на случайном выборе вектора приращения весов по принципу Монте-Карло. Выбор вектора приращения весов Δw_{ji} реализуется поиском в пространстве весов, при этом принимаются только те изменения Δw_{ji} , которые ведут к уменьшению квадратичной ошибки на каждой итерации. Метод достаточно прост по вычислениям, но недостаточно эффективен и медленно сходится.

Метод Бройдена – Флетчера. Метод нелинейной минимизации Бройдена – Флетчера основан на методе Ньютона [36], в котором используется разложение функции ошибок, включающее вторые производные



$$\frac{\partial^2 E(\mathbf{w})}{\partial w_i \partial w_j}, \text{ т.е. матрицу Гессе } \mathbf{H} = \left[H_{ij} = \frac{\partial^2 E(\mathbf{w})}{\partial w_i \partial w_j} \right]:$$

$$E(\mathbf{w}(n) + \Delta \mathbf{w}(n)) \cong E(n) + \mathbf{g}^T(n) \cdot \Delta \mathbf{w}(n) + \frac{1}{2} \Delta \mathbf{w}^T(n) \cdot \mathbf{H}^T, \quad (27)$$

$$\text{где } E(n) \equiv E(\mathbf{w}(n)); \quad \{\mathbf{g}(n)\}_i \equiv \frac{\partial E(n)}{\partial w_i(n)};$$

$$\{H(n)\}_{ij} = \frac{\partial^2 E(\mathbf{w}(n))}{\partial w_i(n) \partial w_j(n)}; \quad \mathbf{g}(n) - \text{градиент ошибки.}$$

Изменение весов $\Delta \mathbf{w}(n)$ вектора параметров персептрона $\mathbf{w}(n)$ на n -й итерации определяется условием

$$\Delta \mathbf{w}(n) = \arg \min_{\Delta \mathbf{w}} \{E(\mathbf{w}(n) + \Delta \mathbf{w}(n))\} \quad (28)$$

Соотношение (28) с учетом (27) эквивалентно выражению (29), которое лежит в основе получения $\Delta \mathbf{w}(n)$ по методу Ньютона:

$$\mathbf{H}(n) \cdot \Delta \mathbf{w}(n) = -\mathbf{g}(n) \quad (29)$$

При решении матричного уравнения (29) относительно $\Delta \mathbf{w}(n)$ необходимо, чтобы матрица Гессе была положительно определенной и хорошо обусловленной. Но эти условия не всегда выполняются на практике.

Модификация Бройдена – Флетчера для метода Ньютона связана с построением положительно определенной матрицы $\tilde{\mathbf{H}}(n)$, которая является аппроксимацией матрицы Гессе на основе конечных разностей градиента функции ошибок [42–46]. Матрицей $\tilde{\mathbf{H}}(n)$ представляется разложение (30) градиента рядом Тейлора, которое в более короткой форме эквивалентно (31):

$$\mathbf{g}(\mathbf{w}(n) + \Delta \mathbf{w}(n)) \cong \mathbf{g}(\mathbf{w}(n)) + \tilde{\mathbf{H}}(n) \cdot \Delta \mathbf{w}(n) \quad (30)$$

$$\tilde{\mathbf{H}}(n) \cdot \Delta \mathbf{w}(n) = \mathbf{y}(n) \quad (31)$$

где через $\mathbf{y}(n) \equiv \mathbf{g}(n+1) - \mathbf{g}(n)$ обозначено приращение градиента на n -й итерации.

В методе Бройдена-Флетчера аппроксимация матрицы Гессе в алгоритме обучения вычисляется итеративно с помощью соотношения (32):

$$\tilde{\mathbf{H}}(n+1) = \tilde{\mathbf{H}}(n) - \frac{1}{\Delta \mathbf{w}^T(n) \cdot \tilde{\mathbf{H}}(n) \cdot \Delta \mathbf{w}(n)} \tilde{\mathbf{H}}(n) \cdot \Delta \mathbf{w}(n) \cdot \Delta \mathbf{w}^T(n) + \frac{1}{\mathbf{y}^T(n) \cdot \Delta \mathbf{w}(n)} \mathbf{y}(n) \cdot \mathbf{y}^T(n) \quad (32)$$

При этом вектор $\mathbf{p}(n)$ направления изменения параметров персептрона $\Delta \mathbf{w}(n)$ определяется соотношением (33), которое по форме аналогично выражению характерному для метода Ньютона [36]:

$$\tilde{\mathbf{H}}(n) \cdot \mathbf{p}(n) = -\mathbf{g}(n) \quad (33)$$

Изменение весов персептрона $\Delta \mathbf{w}(n)$ вдоль этого направления с параметром скорости обучения η определяется соотношением (34):

$$\Delta \mathbf{w}(n) = \eta \cdot \mathbf{p}(n) \quad (34)$$

Таким образом, выражение (32) преобразуется к ви-

ду (35):

$$\tilde{\mathbf{H}}(n+1) = \tilde{\mathbf{H}}(n) + \frac{1}{\mathbf{g}^T(n) \cdot \mathbf{p}(n)} \mathbf{g}(n) \cdot \mathbf{g}^T(n) + \frac{1}{\eta(n) \mathbf{y}(n) \cdot \mathbf{p}(n)} \mathbf{y}(n) \cdot \mathbf{y}^T(n) \quad (35)$$

Выражения (33) – (35) служат основой алгоритма обучения персептрона по методу Бройдена – Флетчера. Алгоритм включает оценку градиента $\mathbf{g}(n)$ функции ошибок $E(\mathbf{w}(n))$, определение вектора направления $\mathbf{p}(n)$ и схему обновления матрицы $\tilde{\mathbf{H}}(n)$ [36].

Метод Бройдена – Флетчера использует не только первую, но и вторую производную функции ошибок (в форме ее разностной аппроксимации) и более точно определяет направление при поиске минимума.

Метод параллельных касательных (метод partan).

Метод реализует движение в направлении, которое параллельно касательной к линии равного уровня поверхности ошибок, найденной градиентным методом для предшествующей точки [36, 47]. Метод использует матрицу Гессе (36) и алгоритм сопряженных направлений $\mathbf{S}_n$ и $\mathbf{S}_{n+1}$ по соотношению (37):

$$\mathbf{H}[E(\mathbf{w}_n)] = \nabla^2 \{E[\mathbf{w}(n)]\},$$

$$\text{т.е. } H_{ij} = \frac{\partial^2 E(\mathbf{w}_n)}{\partial w_i \partial w_j} \quad (36)$$

$$(\mathbf{S}_{n+1})^T \cdot [\nabla^2 \{E[\mathbf{w}(n)]\}] \mathbf{S}_n = 0 \quad (37)$$

Условие (37) – это условие ортогональности вектора направления $\mathbf{S}_{n+1}$ на $(n+1)$ шаге относительно вектора $[\nabla^2 \{E[\mathbf{w}(n)]\}] \mathbf{S}_n$, где $\nabla^2 \{E[\mathbf{w}(n)]\}$ – матрица Гессе (36).

Корректировка вектора весов нейроперсептрона выполняется по соотношению:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \lambda_n \mathbf{S}_n \quad (38)$$

Значение параметра λ , соответствующее минимуму $E(\mathbf{w}_{n+1})$ в направлении вектора $\mathbf{S}_n$, определяется выражением

$$\lambda_n = - \frac{[\nabla E(\mathbf{w}_n)]^T \cdot \mathbf{S}_n}{\mathbf{S}_n^T \cdot [\nabla^2 \{E(\mathbf{w}_n)\}] \cdot \mathbf{S}_n}, \quad (39)$$

где $\nabla E(\mathbf{w}_n)$ – вектор градиента

$$[\nabla E(\mathbf{w}(n))]_i = [\text{grad} \{E(\mathbf{w}_n)\}]_i \equiv \frac{\partial E(\mathbf{w}(n))}{\partial w_i(n)}.$$

Таким образом, метод параллельных касательных требует вычисления гиперплоскости, касательной к функции ошибок, на каждой итерации. Поэтому метод связан с большими вычислительными затратами, особенно в задачах большой размерности. В двухмерном варианте метод основан на движении по направлению, параллельному касательной к линии постоянного уровня функции ошибок, однако при большой размерности аргумента (СВ персептрона) используются касательные гиперплоскости и движение в направлении, параллельном этим гиперплоскостям. Требование вычисления касательных гиперплоскостей и определения $\min_{\mathbf{w}_n} \{E(\mathbf{w}_n)\}$ вдоль направления, параллельного этим гиперплоскостям, является недостатком

метода. При значительной размерности СВ (в функции ошибок персептрона) метод постепенно теряет эффективность, снижая точность и увеличивая длительность обучения.

Выбор метода обучения нейроперсептрона для решаемой задачи. При использовании любого алгоритма обучения важна инициализация СВ и порогов нейронов сети. В качестве начальных значений весов и уровней порога нейронов для создания им равноценных условий подстройки в процессе обучения целесообразно выбирать равномерно распределенные числа. Значения параметров синаптической связи для нейрона i в сети рекомендуется обычно ограничивать интервалом $[-2.4(F_i)^{-1}, 2.4(F_i)^{-1}]$, где F_i – общее количество входов i -го нейрона [48].

Обобщая представленный выше анализ, можно утверждать, что в алгоритмах обучения персептрона для задачи детектирования речевой активности может быть использован один из рассмотренных методов, хотя наиболее совершенными представляются два из них:

- обобщенный метод дельта–дельта (правило Джекобса), который включает адаптивный выбор параметра скорости обучения;
- метод Бройдена–Флетчера, который использует приближение для производных второго порядка (матрицу Гессе) и позволяет определять величину параметра скорости обучения.

Однако только на основе теоретического анализа нельзя сделать окончательный выбор метода обучения персептрона задаче распознавания речевого сигнала, поскольку невозможно учесть поведение алгоритмов в условиях многообразия характеристик сигнала и шума, величин пауз и наличия переходных сегментов. Исходя из этого, для обоснованного выбора метода обучения, на достаточном представительных данных проведена экспериментальная проверка алгоритмов при различных видах шумов и различных отношениях сигнал/шум.

Экспериментальное исследование характеристик алгоритмов обучения нейроперсептрона. В целях объективной оценки преимуществ алгоритмов обучения для детектирования речевой активности проведен компьютерный эксперимент, в котором на речевой сигнал, включающий интервалы активности и микропаузы, наложены различные виды шума:

- белый шум, имеющий равномерный спектр мощности в звуковом диапазоне частот;
- окрашенный шум с постепенным спадом спектра при росте частоты (для розового шума – 6 дБ/октаву), представителем которого может служить шум пылесоса или шум льющейся воды;
- музыкальный фон;
- узкополосный шум, состоящий из детерминированных или случайных компонент, постоянных по своему положению и амплитуде (шум вентилятора).

Эксперимент проведен с использованием нейросетевого эмулятора Essence 1.0 [49], выполненного на языке Java и позволяющего программно моделировать многослойный нейроперсептрон с любым количеством слоев, устанавливать отдельно количество и вид активационной функции нейронов для каждого слоя, выполнять все перечисленные

алгоритмы обучения и использовать различные виды ошибок.

Данными служили оцифрованные записи речевого сигнала и сопутствующего шума с частотой дискретизации 22.5 кГц. На каждом сегменте записи длительностью 20 мс выполнено вейвлет–разложение в соответствии с персептуальной моделью, вычислена средняя мощность вейвлет–коэффициентов в каждой области разложения и полученные векторы мощности вейвлет–коэффициентов нормированы.

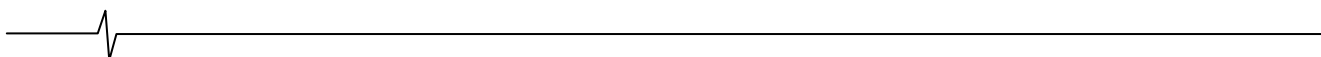
На основе анализируемого зашумленного речевого сигнала получено 2000 векторов, которые составили два множества (обучающее и тестовое) по 1000 векторов. Использован супервизорный принцип обучения (с учителем), при котором сети сообщается желаемый отклик на каждый входной обучающий вектор. Программа представляет динамику изменения ошибки в процессе обучения и на тестирующей выборке. При сходимости обучения ошибка почти не изменяется, но если она недостаточно мала, то сеть усложнялась и повторялось ее обучение и тестирование. Напротив, если ошибка по обучающей выборке стремится к нулю, а при тестировании сохраняет недопустимо большое значение, то размер сети уменьшался.

После ряда экспериментов с изменением размера сети для исследования различных алгоритмов обучения принята двухслойная сеть с числом нейронов десять и один (10-1) и с логистической активационной функцией нейронов. Для всех алгоритмов при равном времени обучения (2 мин) зафиксировано число циклов обучения и процент ошибочных решений по обучающей и тестовой выборкам. Эксперимент проведен для всех видов шума и отношениях среднеквадратического отклонения шума к сигналу $\zeta = 0.02, 0.1$ и 0.3 .

Результаты компьютерного эксперимента, позволившие сопоставить различные алгоритмы обучения, показывает таблица 3, где приведено число циклов обучения и процент ошибочных решений при обучении и тестировании для нескольких уровней сложности нейронной сети, типов и уровней сопутствующего шума. Число циклов различается в зависимости от алгоритма и определяется его вычислительной сложностью. Для всех методов обучения процент ошибочных решений на обучающей выборке меньше, чем на тестовой выборке. Однако это отличие не превышает 20 %, что подтверждает адекватность выбранной модели нейронной сети.

При сопоставлении алгоритмов по оценке вероятности ошибочных решений наилучшие характеристики получены для метода Бройдена–Флетчера. Вероятность ошибочных решений зависит от интенсивности шума: для $\zeta = 0.02$ вероятность ошибочных решений по тестовой выборке не превышает 3.7 %, для $\zeta = 0.1$ – 4.3 %, а для $\zeta = 0.3$ – 5.5 %. Вероятность ошибочных решений для узкополосного шума и музыкального фона выше, поскольку спектральный состав шума меньше отличается от речевого сигнала.

Различаются ошибочные решения первого и второго рода [50, 51]. Пусть ω_1 – событие, состоящее в том, что текущий сегмент или фрейм содержит речевой сигнал; ω_2 – событие, состоящее в том, что текущий фрейм содержит шум; Ω_1 – решение о том, что текущий фрейм содержит речевой сигнал; Ω_2 – решение о том, что текущий фрейм содержит шум. Тогда вероятность ошибочного обнаруже-



ния речевого сигнала во время паузы определяется соотношением:

$$P(\Omega_1 / \omega_2) = \frac{P(\Omega_1, \omega_2)}{P(\omega_2)} \quad (40)$$

Вероятность ошибочного обнаружения паузы при наличии речевого сигнала – соотношением:

$$P(\Omega_2 / \omega_1) = \frac{P(\Omega_2, \omega_1)}{P(\omega_1)} \quad (41)$$

Компьютерный эксперимент показал, что вероятность ошибочного обнаружения паузы при наличии речевого сигнала ниже, чем вероятность ошибочного обнаружения речевого сигнала.

Таблица 3

Сравнение методов обучения персептрона

Вид шума	Относи- тельный вклад шума	Метод обучения	Обучение			Тести- рование
			Число циклов	Время, мин	Ошибка %	Ошибка %
1	2	3	4	5	6	7
Белый шум	0.02	Случайный поиск	500	5	10	13
		Сопряжен. градиенты	800	3.2	3.5	5.7
		Partan	600	3.3	4	4.7
		Дельта-дельта	600	3.3	4.5	5.2
		Обобщ. дельта-дельта	650	4	2.5	3.8
		Бройдена – Флетчера	700	2	2.7	3.1
	0.1	Случайный поиск	500	6.5	12	15
		Сопряжен. градиенты	900	4.5	5	8
		Partan	700	4	6.5	6.9
		Дельта-дельта	700	4	6.5	7
		Обобщ. дельта-дельта	800	4.4	5	5.3
		Бройдена – Флетчера	800	2.2	3.4	3.6
Белый шум	0.3	Случайный поиск	900	7	14	15
		Сопряжен. градиенты	900	5	7	8.2
		Partan	1000	4.8	9.5	9.3
		Дельта-дельта	700	4	9	5.5
		Обобщ. дельта-дельта	900	45	7	7.3
		Бройдена – Флетчера	1000	2.5	4.5	4.7
Окрашенный шум	0.02	Случайный поиск	600	5	11	12
		Сопряжен. градиенты	800	3.3	3.7	6.0
		Partan	600	3.5	4.1	4.9
		Дельта-дельта	600	3.5	4.6	5.5
		Обобщ. дельта-дельта	700	4.5	2.4	4
		Бройдена – Флетчера	800	2.1	2.6	3.3
	0.1	Случайный поиск	500	6.4	12.5	13
		Сопряжен. градиенты	850	4.6	4.8	4.8
		Partan	800	4.3	6	6.8
		Дельта-дельта	600	4.1	7	7
		Обобщ. дельта-дельта	850	4	4.2	4.2
		Бройдена – Флетчера	900	2.0	3.2	3.4
	0.3	Случайный поиск	600	5	14.5	15
		Сопряжен. градиенты	800	4	6.5	6.8
		Partan	650	3.5	8	8.5
		Дельта-дельта	700	3.0	7	8.3
		Обобщ. дельта-дельта	900	3.6	6.5	6
		Бройдена – Флетчера	800	2	4.8	5
Музыкальный фон	0.02	Случайный поиск	500	4	10	10.5
		Сопряжен. градиенты	750	3.1	3.8	4.0
		Partan	650	3.5	4	4.2
		Дельта-дельта	600	3.3	4.5	4.6
		Обобщ. дельта-дельта	750	4.0	2.6	2.8
		Бройдена – Флетчера	900	2.8	2.9	3.0
	0.1	Случайный поиск	400	7	12	12.5
		Сопряжен. градиенты	800	5	5	5.2
		Partan	850	4	5.8	6
		Дельта-дельта	550	4.2	7	7.1
		Обобщ. дельта-дельта	800	3.9	4.1	4.2
		Бройдена – Флетчера	1000	2.3	3.1	3.3

Таблица 3 (продолжение 1)

1	2	3	4	5	6	7
	0.3	Случайный поиск	600	4.5	14	15.2
		Сопряжен. градиенты	700	4	6.7	6.9
		Partan	600	3.2	8	8.2
		Дельта-дельта	650	3.2	7.3	8
		Обобщ. дельта-дельта	800	3.5	5.3	5.6
		Бройдена – Флетчера	900	1.8	6	5.5
Узкополосный шум	0.2	Случайный поиск	500	5	10.5	10
		Сопряжен. градиенты	700	3.2	4.5	4.7
		Partan	600	3.3	5	5.4
		Дельта-дельта	600	3.6	5.8	6.2
		Обобщ. дельта-дельта	600	4	2.6	2.8
		Бройдена – Флетчера	600	2.6	3.5	3.7
Узкополосный шум	0.1	Случайный поиск	600	4.5	12	11.8
		Сопряжен. градиенты	700	3.8	6.0	6.4
		Partan	550	3.5	6	6.6
		Дельта-дельта	700	3.8	7.0	7.3
		Обобщ. дельта-дельта	800	4.2	3.5	3.8
		Бройдена – Флетчера	800	4.5	4.1	4.3
Узкополосный шум	0.3	Случайный поиск	750	4	14	15
		Сопряжен. градиенты	900	5	8	9
		Partan	800	4.3	8.2	8.1
		Дельта-дельта	900	4.5	9	9.5
		Обобщ. дельта-дельта	1000	5	4.9	5.3
		Бройдена – Флетчера	1000	2.8	3.2	3.6

Выбор структуры персептрона. Выбор структуры многослойного персептрона (МСП) важен, поскольку повышенная сложность сети приводит к увеличению объема вычислений и длительности обучения, возникновению неустойчивости и переобученности, которая ухудшает показатели точности на материале тестирования. Чрезмерное упрощение в свою очередь не позволяет получить хорошее качество детектирования свойств сигнала. Поэтому построение нейронной сети для системы детектирования состояния процесса связано с выбором минимальной конфигурации, обеспечивающей хорошее функционирование.

Для сети минимального размера маловероятно обучиться несущественным деталям в обучающих данных, поэтому она может давать лучшие результаты на материале тестирования. Способ реализации такого подхода – это выбор сети малой сложности и наращивание ее до получения приемлемых показателей алгоритма обучения и правильности работы на тестовом материале. Другой подход состоит в начальном выборе сети достаточно сложной структуры, обеспечивающей приемлемую ошибку обучения, после чего производится упрощение структуры сети путем удаления избыточных нейронов. Дополнительным ресурсом является метод регуляризации структуры персептрона, в котором вместо простого критерия среднеквадратической ошибки функция риска учитывает также меру сложности структуры персептрона.

Выбор структуры нейронной сети осуществлен на основе совместного использования методов упрощения и наращивания и метода регуляризации.

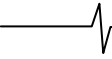
Для метода наращивания в ходе проведения серии циклов обучения, основываясь на критерии ошибки (12), определялось место размещения в сети дополнительного нейрона. Одновременно из сети удалялся нейрон, веса которого флуктуировали после сходимости обучения, т.к. этот нейрон не является существенным для задачи обуче-

ния. Результаты моделирования получены на ЭВМ с использованием программы Essence 1.0 для речевого сигнала с белым шумом при отношении среднеквадратического отклонения шума и сигнала 0.1 и 0.3. Метод наращивания сложности МСП проанализирован для структур: 3-1; 5-1; 5-1-1; 5-2-1; 5-3-1 и двух лучших алгоритмов обучения: Бройдена-Флетчера и обобщенного правила дельта-дельта (табл.4). Введение третьего нейрона во втором скрытом слое МСП несущественно меняет ошибку, но обнаруживает флуктуации вектора весов дополнительного нейрона в конце цикла обучения. Поэтому метод указывает на предпочтительность структур 5-2-1.

Таблица 4
Оптимальная структура МСП по методу наращивания

Структура МСП	Отношение шум/сигнал	Доля ошибочных решений персептрона (%)	
		Обобщенное правило дельта-дельта	Алгоритм Бройдена-Флетчера
3-1	0.1	8.1	6.5
	0.3	9.7	7.0
5-1	0.1	6.5	5.0
	0.3	8.3	5.5
5-1-1	0.1	4.9	4.0
	0.3	6.9	4.5
5-2-1	0.1	3.7	3.2
	0.3	5.5	4.0
5-3-1	0.1	3.9	3.4
	0.3	5.8	4.2

Метод упрощения, где после каждого цикла обучения удалялся нейрон, отсутствие которого дает минимальное увеличение ошибки обучения, применен к структурам персептрона 10-3-1; 10-2-1; 8-2-1; 5-2-1 (табл.5). Результаты анализа согласуются с выводом, полученным по методу наращивания.



Метод регуляризации [33] вместо критерия среднеквадратической ошибки использует функцию риска $R(\mathbf{w})$:

$$R(\mathbf{w}) = E(\mathbf{w}) + \lambda \cdot E_c(\mathbf{w}), \quad (42)$$

где $E(\mathbf{w})$ – среднеквадратическая ошибка, $E_c(\mathbf{w})$ – мера сложности структуры персептрона, а параметр λ определяет относительную значимость компоненты $E_c(\mathbf{w})$ и служит параметром регуляризации (поскольку форма критерия $R(\mathbf{w})$ опирается на теорию регуляризации А.Н. Тихонова). В качестве меры сложности структуры персептрона использован квадрат нормы вектора синаптических весов МСП:

$$E_c(\mathbf{w}) = \|\mathbf{w}\|^2 = \sum_{i \in C} w_i^2, \quad (43)$$

где C – множество индексов всех весов сети. Эксперимент проведен на структурах МСП 5-2-1 и 5-3-1 с изменением параметра регуляризации от 0 (в исходном состоянии сети) до 0.01 с шагом 0.001 и показал, что один из нейронов второго скрытого слоя в 5-3-1 имеет 1.5 %-й вклад в $E_c(\mathbf{w})$ и может считаться несущественным. Это подтверждает результат, полученный методом наращивания сложности персептрона.

Таблица 5
Результаты определения оптимальной структуры МСП методом упрощения

Структура персептрона	Отношение шум/сигнал	Доля ошибочных решений персептрона (%)	
		Обобщенное правило дельта-дельта	Алгоритм Бройдена-Флетчера
10-3-1	0.1	2.1	2.3
	0.3	4.2	3.1
10-2-1	0.1	2.3	2.5
	0.3	4.5	3.6
8-2-1	0.1	2.9	2.8
	0.3	5.0	3.8
5-2-1	0.1	3.7	3.2
	0.3	5.5	4.0

Использование главных компонент сигнала для обучения распознаванию. Дальнейшее улучшение алгоритма детектирования речевой активности достигнуто уменьшением размерности входного вектора персептрона, которое позволило уменьшить число синаптических весов, уменьшить объем обучающей выборки и ускорить процесс обучения. Уменьшение размерности входных векторов при условии сохранения содержащейся в них информации о речевом сигнале основано на методе главных компонент [52, 53].

Входной вектор МСП $\mathbf{x}_j = [x_{1j}, x_{2j}, \dots, x_{pj}]$ имеет $p = 22$ компоненты, число которых соответствует персептальной модели. Идея метода состоит в получении p линейных комбинаций из компонент вектора – главных компонент, каждая из которых линейно независима от других. Главная компонента $\mathbf{y}$ выражается соотношением:

$$\mathbf{y} = \mathbf{X} \hat{\boldsymbol{\beta}}, \quad (44)$$

где матрица $\mathbf{X}$ имеет вид:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ x_{1N} & x_{2N} & \dots & x_{pN} \end{bmatrix}.$$

Строками матрицы $\mathbf{X}$ служат p -мерные обучающие векторы, поступающие на вход персептрона, ее размерность равна $N \times p$ (N – объем обучающей выборки). Вектор

$$\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p]^T -$$

вектор неизвестных коэффициентов. Каждый такой вектор по (44) определяет соответствующую ему главную компоненту.

Выборочная дисперсия $\mathbf{y}$ определяется выражением:

$$\text{var}(\mathbf{y}) = \hat{\boldsymbol{\beta}}^T \mathbf{S} \hat{\boldsymbol{\beta}}, \quad (45)$$

где $\mathbf{S}$ – выборочная ковариационная матрица обучающих векторов, составляющих матрицу $\mathbf{X}$:

$$\mathbf{S} = \frac{1}{N-1} \begin{bmatrix} \sum_{i=1}^N x_{1i}^2 & \sum_{i=1}^N x_{1i}x_{2i} & \dots & \sum_{i=1}^N x_{1i}x_{pi} \\ \sum_{i=1}^N x_{2i}x_{1i} & \sum_{i=1}^N x_{2i}^2 & \dots & \sum_{i=1}^N x_{2i}x_{pi} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^N x_{pi}x_{1i} & \sum_{i=1}^N x_{pi}x_{2i} & \dots & \sum_{i=1}^N x_{pi}^2 \end{bmatrix} \quad (46)$$

Главные компоненты, начиная с первой $\mathbf{y}_1$, определяются по критерию максимальной дисперсии $\hat{\boldsymbol{\beta}}^T \mathbf{S} \hat{\boldsymbol{\beta}}$ при условии, что вектор коэффициентов $\hat{\boldsymbol{\beta}}$ нормирован, т.е. $\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}} = 1$. Решение условной максимизации использует функцию $\varphi = \hat{\boldsymbol{\beta}}^T \mathbf{S} \hat{\boldsymbol{\beta}} - \hat{\lambda}(\hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\beta}} - 1)$ ($\hat{\lambda}$ – множитель Лагранжа), градиент которой $d\varphi/d\hat{\boldsymbol{\beta}} = 2\mathbf{S}\hat{\boldsymbol{\beta}} - 2\hat{\lambda}\hat{\boldsymbol{\beta}}$ после приравнивания к нулю дает уравнение $(\mathbf{S} - \hat{\lambda}\mathbf{I})\hat{\boldsymbol{\beta}} = 0$ для определения компонент вектора $\hat{\boldsymbol{\beta}}$. Это матричное уравнение имеет решение, если равен нулю определитель $|\mathbf{S} - \hat{\lambda}\mathbf{I}| = 0$. Определитель представляет собой алгебраическое уравнение степени p относительно $\hat{\lambda}$ и позволяет найти p корней, являющихся характеристическими числами $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$ ковариационной матрицы $\mathbf{S}$. Наибольшее $\hat{\lambda}_1$ максимизирует дисперсию $\mathbf{y}_1$, а решение системы $(\mathbf{S} - \hat{\lambda}_1 \mathbf{I})\hat{\boldsymbol{\beta}} = 0$ дает вектор $\hat{\boldsymbol{\beta}}_1$, определяющий первую главную компоненту.

Следующие по величине характеристические числа в качестве решения системы $(\mathbf{S} - \hat{\lambda}_1 \mathbf{I})\hat{\boldsymbol{\beta}} = 0$ дают векторы $\hat{\boldsymbol{\beta}}_2$, которые определяют остальные главные компоненты по соотношению (44). Таким образом, для p -мерных данных первая главная компонента $\mathbf{y}_1 = \mathbf{X} \cdot \hat{\boldsymbol{\beta}}_1$ представляет собой линейную комбинацию p компонент с коэффициентами, равными нормированному собственному вектору корреляционной матрицы $\mathbf{S}$, соответствующему максимальному характеристическому числу этой матрицы. Аналогично остальные главные компоненты соответствуют следующим по величине характеристическим числам матрицы $\mathbf{S}$. Главные компоненты линейно независимы, и дисперсия каждой компоненты равна соответствующему характеристическому числу матрицы $\mathbf{S}$.

Общая дисперсия p компонент равна следу (tr) выборочной ковариационной матрицы:

$$tr(\mathbf{S}) \equiv var(y_1) + var(y_2) + \dots + var(y_p).$$

Относительная значимость j -й главной компоненты определяется соотношением

$$\frac{var(y_j)}{\sum_{j=1}^p var(y_j)} = \frac{\hat{\lambda}_j}{tr(\mathbf{S})} = \frac{\hat{\lambda}_j}{\sum_{j=1}^p \hat{\lambda}_j} \quad (47)$$

Для уменьшения размерности выполнен анализ и выбраны q наиболее значимых компонент $q \leq p$, которые определяют η % от общей дисперсии всех p компонент. Для $\eta = 95 \dots 98$ % вклад оставшихся $p - q$ компонент может считаться несущественным.

При проведении исследования программа Essence 1.0 дополнена методом главных компонент. Исходными данными служили 100 векторов, представляющих нормированную мощность коэффициентов вейвлет-разложения речевого сигнала в соответствии с перцептуальной моделью.

Отношение сигнал/шум варьировалось от 3 до 50. Для каждого из четырех уровней шума речевой сигнал, преобразованный вейвлет-пакетом и имеющий первоначально 22 компоненты, обработан программой главных компонент, в результате чего получены 9 компонент, содержащих не менее 97.5 % энергии исследуемого сигнала. При этом допустимый уровень потери энергии составил не более $\eta = 2.5$ %. Использование главных компонент в структуре перцептрона показано на рисунке. Обучение перцептрона проведено с использованием алгоритмов Бройдена – Флетчера и обобщенного правила дельта-дельта, т. к. они дали лучшие результаты обучения на исходных обучающих векторах. При уменьшении отношения сигнал/шум (в исследованном диапазоне 3 – 50) ошибка обучения возрастает на 60 % для метода Бройдена – Флетчера и вдвое – для обобщенного правила дельта-дельта Джекобса. Для белого и окрашенного шумов метод обобщенного правила дельта-дельта имеет чуть меньшую ошибку обучения (на 0.1 %) при малом отношении шум/сигнал (до 0.05) и уступает при больших уровнях шума.

Для условий узкополосного шума и музыкального фона метод обобщенного правила дельта-дельта дает меньшую ошибку обучения и тестирования на всем диапазоне уровней шума от 0.5 до 0.2 %.

Для проверки этой закономерности проанализированы алгоритмы обучения по методу сопряженных градиентов и методу параллельных касательных (метод *partan*), которые в аналогичных условиях давали в среднем на 1 % большую ошибку обучения, чем первые два метода. Для белого и окрашенного шумов метод параллельных касательных уступает методу сопряженных градиентов – ошибка на 1 % больше во всем диапазоне отношений шум/речь. Для музыкального фона это различие несколько меньше, а для узкополосного шума метод параллельных касательных дает уже на 0.5 % меньшую ошибку обучения.

Методы Бройдена – Флетчера и обобщенное правило дельта-дельта дают близкие показатели относительной ошибки и количество циклов обучения.

Метод главных компонент дает возможность снизить количество циклов на 30–35 % (в зависимости от метода и ви-

да шума) и уменьшить время обучения на 40–50 %. В то же время ошибка обучения увеличивается на 0.4–0.6 %. Для узкополосного шума и музыкального фона это снижение меньше, чем для белого и окрашенного шумов, во всем диапазоне уровней шума.

Применение метода главных компонент позволило:

- уменьшить размерность входного вектора нейросептрона с 22 до 9 компонент;
- уменьшить число оцениваемых в процессе обучения свободных параметров сети (синаптических весов) с 122 до 52 – для нейронной сети 5-2-1;
- соответственно увеличить скорость обучения и уменьшить объем обучающей выборки.

Заключение. Для детектирования речевой активности предложен метод, включающий вейвлет-преобразование сигнала в соответствии с перцептуальной моделью и нейросетевой алгоритм (см. раздел I). Аппроксимация перцептуальной модели реализована на основе вейвлет-пакета, разделяющего частотный диапазон сигнала на 22 субполосы.

Принятие решения о наличии речевой активности на анализируемом сегменте осуществлено на основе нейронной сети в форме многослойного перцептрона (МСП). На входы МСП поступает нормированный вектор с компонентами, равными мощности вейвлет-коэффициентов в субполосах разложения сигнала. Исследование показало, что для обучения МСП целесообразно использовать метод Бройдена – Флетчера, который по результатам компьютерного эксперимента имеет наилучшие характеристики по скорости обучения, величине ошибок (на обучающей и тестовой выборке) и по вероятности ошибочной классификации сегмента сигнала.

Проведены экспериментальные исследования по регуляризации структуры перцептрона, основанные на методах наращивания, упрощения и регуляризации, которые показали, что оптимальной для решения данной задачи является структура перцептрона 5-2-1.

Применение метода главных компонент позволило уменьшить размерность входного вектора нейронной сети, уменьшить число оцениваемых в процессе обучения синаптических весов и увеличить скорость обучения при уменьшенном объеме обучающей выборки.

Литература

1. Кратиров Д.В., Меркушева А.В. Алгоритм, основанный на вейвлет-преобразовании и нейронной сети, для бесконтактного измерения параметров газожидкостного потока / Сборник докладов Международной конференции «Датчики и системы». СПб.: Изд-во СПбГТУ, 2002. Т. 3. С.51–55.
2. Daubechies I. Painless No orthogonal Expansions / Journal of Mathematical. Physics, 1986, v. 27, p. 1271 – 1283.
3. Villemois L.F. Energy Moments in Time and Frequency for 2-Scale Equations and Wavelets / SIAM Journal of Mathematical Analysis, 1992, v. 23, p. 1119 – 1153.
4. Рабинер Л.П., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. – 492 с.
5. Allen J.B., Rabiner L.R. A Unified Approach to Short-Time Fourier Analyses and Synthesis / Proceedings of IEEE, 1977, v. 65, n. 11, p. 1558.
6. Portnoff M.R. Time-Frequency Representation of Digital Signals and Systems Based on Short-Time Fourier Analyses / IEEE

Transactions on Signal Processing, 1980, v. 28, n. 2, p. 55.

7. Исмаилов Ш.Ю., Меркушева А.В. Нейросетевой алгоритм на вейвлет-преобразовании нестационарного сигнала в ИИС / Сб. докладов Международной конференции по мягким вычислениям и измерениям SCM-2001. СПб., 2001, том 1, с. 251 – 256.
8. Малыгина Г.Ф., Меркушева А.В. Вейвлет-фильтрация нестационарного сигнала с адаптацией на основе нейронной сети / Сб. докладов Международной конференции по мягким вычислениям и измерениям, SCM-2001. СПб: ГЭТУ (ЛЭТИ), 2001, т. 1, с. 239 – 242.
9. Lang M., Guo H. Noise Reduction Using Undecimated Discrete Wavelet Transform / IEEE Signal Processing Letters, 1996, v. 3, n1, p. 8 – 10.
10. Atal B.S. Optimal Noise Specter Form / IEEE Transactions on Signal Processing, 1979, n. 6, p. 247 – 254.
11. Fletcher N. Auditory Patterns / Review of Modern Physics, 1940, p. 47–65.
12. Zwicker E., Fastl H. Psychoacoustics, Facts and Models. Berlin: Springer-Verlag, 1984. – 420 c.
13. Scharf B. Critical Bands // Foundation of Modern Auditory Theory / Ed. Tobias J. N.Y.: Acad. Press, 1970, p. 159–222.
14. Schroeder M.R., Atlas B.C., Hall J.L. Optimizing Digital Speech Coders by Exploiting the Masking Property / Journal of Audio Engineering Society, 1995, v. 43, n 11, p. 914.
15. Zelinski R., Noll P. Adaptive Bit Allocation with Thin Structure (on Formants) of Optimal Distortion or Noise Specter / IEEE Transactions on Signal Processing, 1977, n. 8, p. 299 – 309.
16. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов (перев. с англ.). М.: Мир, 1978. – 848 с.
17. Малыгина Г.Ф., Меркушева А.В. Детектирование речевого сигнала и фильтрация с адаптивным порогом / Сборник трудов факультета техн. кибернетики “Микропроцессорные средства измерений”. СПб: СПбГТУ, 2001, Вып. 2. С. 26–35.
18. Berger J, Coifman R.R., Goldberg M.J. Removing of Noise from Music Using Local Trigonometric Bases and Wavelet-Packets / Journal of Audio Eng. Society, 1994, v. 42, n. 9, p. 808.
19. Chui C.K., Li C. Nonorthogonal Wavelet Packets / SIAM Journal of Mathematical Analysis, 1993, v. 24, p. 712 – 738.
20. Coifman R.R., Wickerhauser M.L. Entropy-Based Algorithms for Best Bases Selection / IEEE Transactions on Information Theory, 1992, v. 38, p. 713–718.
21. Devis G.M., Mallat S., Zhang Z. Adaptive Time-Frequency Decompositions / Optical Engineering, 1994, v. 33, n. 7, p. 2183.
22. Feichtinger H.G. Irregular Sampling Theories and Series Expansions of Band-Limited Functions // SIAM Journal of Mathematical Analysis. 1992. V. 23. P. 530.
23. Ramchadran K., Vetterli M., Herley C. Wavelets, sub-band coding and best bases / Proceedings of IEEE, 1996, v. 84, n. 4, p. 353.
24. Уоссермен Ф. Нейрокомпьютер и его применение (перев. с англ. под ред. А.И. Галушкина). М.: Изд-во Мир, 1992 – 236 с.
25. Цыганков В.Д. Нейрокомпьютер и его применение. М., 1993. – 117 с.
26. Drodle K.W. Unconstrained Optimization / Numerical Analyses, Ed. D. Jackobs,. London: Academic Press, 1977. p. 229 – 388.
27. Russo A.R. Tutorial n. 8 / Proceedings of IEEE Conference on Neural Networks. Washington, 1990, p. 12.
28. Cybenko G. Approximation by Superposition of a Sigmoidal Function / Mathematics of Control, Signals and Systems, 1989, n. 2, p. 303 – 314.
29. Battiti R. First and Second Order Methods for Learning: Between Steepest Descent and Newtons' Method / Neural Computations, 1992, n. 4, p. 141 – 166
30. Baum E.B. Neural Net Algorithms that Learn in Polynoial Time from Examples and Qeries / IEEE Transactions on Neural Networks, 1991, n. 2, p. 5–19.
31. Baum E.B., Wilczek F. Supervised learning of probability distributions by neural networks / Ed. D.Z. Anderson. N.Y.: American Institute of Physics, 1988, p. 52–61.
32. Gallant A.R., White H. / Neural Networks, 1990, n. 5, p. 129–138.
33. Hinton G.E., Nowban S.J. How Learning Can Guide Evolution / Complex systems, 1987, n. 1, p. 495–512.
34. Hopfield J.J. The Effectiveness of Analogue Neural Network.Hardware / Networks, 1990, n. 1, p. 27–40.
35. Saarinen S. et al. Neural networks, back-propagation and automatic differentiation / Automatic Differentiation of Algorithm: Theory, Implementation and Application / Eds. Grievank A., Corless G.F. Philadelphia, MA: SIAM, 1991, p. 31–42.
36. Гилл Ф., Мюррей У., Райт М. Прикладная оптимизация. М.: Мир, 1985 - 209 с.
37. Химмельблау Р. Прикладное нелинейное программирование. М: Мир, 1975, с. 98.
38. Ariel M. Nonlinear Programming: Analyses and Methods. N.Y.: Prentice Hall, 1976.
39. Dorny C.N. A Vector Space Approach to Models and Optimization. N.Y.: Wiley (Interscience Publishing), 1975. - 289 p.
40. Hestenes M.G. Conjugate Direction Method in Optimization. Berlin – Heidelberg – N.Y., 1980 -48 p.
41. Jackobs R.A. Increased Rates of Convergence through Learning Rate Adaptation / Neural Networks, 1989, v. 1, p. 295–307.
42. Broyden C.G. A Class of Methods for Solving Nonlinear Simultaneous Equations / Mathematics of Computation, 1965, n. 19, p. 577–593.
43. Broyden C.G. Quasi-Newton Methods and their Application to Function Minimization / Mathematics of Computation, 1967, n. 21, p. 368–381.
44. Broyden C.G. The Convergence of Minimization Algorithms / Journal of Institute of Mathematical Applications, 1970, n. 6, p. 76–90.
45. Broyden C.G., Dennis J.E., Moro J.J. On the Local and Superlinear Convergence of Quasi-Newton Methods / Journal of Institute of Mathematical Applications, 1973, n. 12, p. 223–245.
46. Shanno D.F. Conditioning of quasi-Newton methods for function initialization / Mathematics of Computation, 1970, n. 24, p. 647–657.
47. Shah B.V., Buchler R.J., Kempthorne O. / Journal Society of Industrial and Applied Mathematics, 1964, v. 12, p. 74.
48. Haykin S. Neural Networks. A Comprehensive Foundation. N.Y.: Prentice Hall, 1994. 680 p.
49. Власов Л.В., Малыгина Г.Ф., Тархов Д.А. Нейронный эмулятор «ESSENCE» / Датчики и системы: Сборник докладов Международной конференции. СПб.: Изд-во СПбГТУ, 2002, т. 3, с. 149–153.
50. Загруднинов Г. М. Достоверность автоматизированного контроля. Казань: Изд-во Казанского университета, 1980. - 280 с.
51. Любатов Ю.В. Теоретические основы моделирования цифровых систем. М.: МАИ, 1989. - 77 с.
52. Болч Б., Хуань К. Дж. Многомерные статистические методы для экономики. М.: Статистика, 1979. - 316 с.
53. Кендал М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды (Пер.с англ. под ред. А.Н. Колмогорова и Ю.В. Прохорова). М.: Изд-во Наука, 1976. - 736 с.

ВЫБОР СТРАТЕГИИ УПРАВЛЕНИЯ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИМИ И ОПЫТНО-КОНСТРУКТОРСКИМИ РАБОТАМИ

Н.Н. Вилкова

Постановка проблемы

В условиях ограничения бюджетных ассигнований на проведение НИОКР резко возрастает роль и значимость рационального управления научными исследованиями и разработками [2-4].

Управление процессами создания образцов техники осуществляется на двух уровнях. Первый уровень (макроуправление) соответствует этапам перспективного планирования наукоемких разработок, формирования годового плана, размещения заказов на создание образцов техники и обеспечения их своевременного и качественного выполнения научно-исследовательскими и производственными подразделениями. Макроуправление является функцией научно-технического подразделения института. Второй уровень (локальное управление) осуществляется в процессе непосредственного выполнения заказов исполнителями и реализуется их руководством, плано-экономическими и финансово-экономическими подразделениями.

При этом локальное управление характеризуется значительной степенью неопределенности в выборе направлений разработок и распределении ресурсов между исполнителями конкретных работ. Это приводит к тому, что сроки выполнения НИОКР могут превышать плановые на 20-30%, а итоговая стоимость выполняемых работ может превосходить ориентировочную на 15 и более процентов.

В связи с изложенным, представляется актуальной и практически значимой проблема определения такой стратегии управления НИОКР, при которой минимизируются

Рассмотрены факторы повышения эффективности управления НИОКР. Представлена формальная модель описания процесса реализации НИОКР и обоснованы основные ситуационные стратегии управления их реализацией.

время и стоимость разработки техники с учётом затрат на автоматизацию [3 - 5]. Данная статья посвящена формальному описанию процесса реализации работ и выбору стратегии управления НИОКР.

Классификация факторов повышения эффективности управления НИОКР

Систематизация представлений о факторах, определяющих эффективность использования ресурсов в НИОКР, необходима для решения задач управления научно-производственной деятельностью: планирования, технико-экономических обоснований, оценки деятельности научно-исследовательской организации (НИО) и т.п. При этом классификация должна обеспечивать, с одной стороны, возможность выделения однородных групп факторов (для решения задач типа «затраты – результаты»), а, с другой стороны, выявления взаимосвязей между факторами (задачи синтеза, получение интегральных оценок и т.п.).

На рис.1 представлена схема факторов повышения эффективности управления ресурсами в НИО. Прежде всего, следует выделить факторы, влияние которых реализуется в рамках научной организации как самостоятельного объекта управления. Кроме того, большое значение имеют внешние факторы, влияние которых реализуется на уровне федеральных органов исполнительной власти. Такое деление, в известном смысле, условно, т.к. имеется связь между внешними и внутренними факторами.

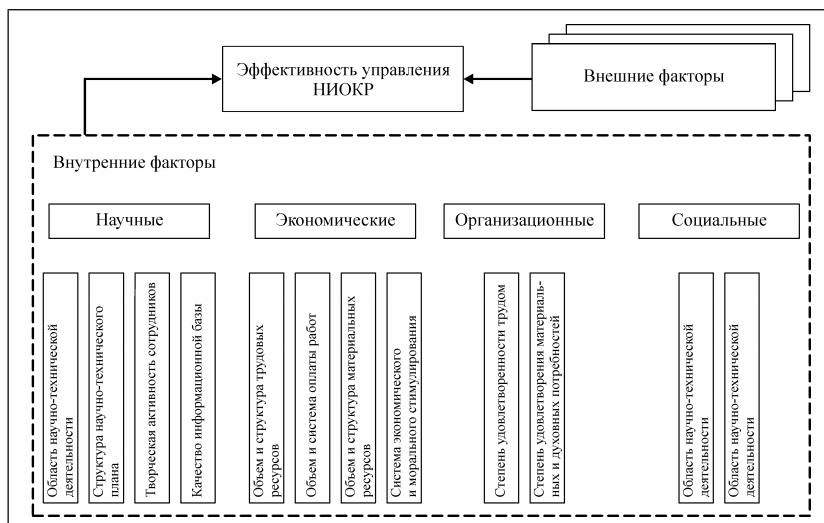
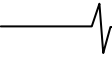


Рис. 1. Схема факторов повышения эффективности управления НИОКР



Факторы, действие которых реализуется в рамках НИО, можно разделить на четыре группы в последовательности, представленной на рис. 1.

При этом первые две группы (научные и экономические факторы) являются определяющими. Если рассматривать НИО как макроэкономическую систему [2, 3], то научные и экономические факторы действуют на входе этой системы, а организационные и социальные определяют динамическое состояние системы, выходами которой являются результаты – качество и эффективность научно – технической продукции.

В связи с развитием новых тематических направлений и сложностью разработки новой техники усложняется и возрастает масштаб исследований, что в условиях ограниченных ресурсов, выделяемых на развитие науки, приводит в настоящее время к повышению роли экономических факторов.

Таким образом, состояние и развитие уровня управления в НИО с использованием современных информационных технологий являются важнейшими факторами повышения эффективности выполнения НИОКР.

Формальное представление процесса реализации НИОКР

Одной из важнейших задач научной организации является управление НИОКР и ресурсами, включающее в себя постановку, организацию выполнения, приостановление, прекращение и выполнение работ [5].

Под ресурсом $R \in \mathcal{R}$ понимается любой объект, который необходим для выполнения НИОКР $P \in \Pi$ и освобождается с её завершением (автоматизированное рабочее место, средство измерений, станок, технологическое оборудование, вычислительные ресурсы и т.п.), либо прекращает своё существование и требует восполнения (материалы, комплектующие изделия и т.п.).

Согласно введенному определению выполняемые работы и используемые ресурсы характеризуются множеством состояний, т.е.

$$\forall i \exists S_{P_i} = \{S_{P_i}^1, S_{P_i}^2, \dots, S_{P_i}^{N_{P_i}}\}, i = \overline{1, N_P}; \quad (1)$$

$$\forall j \exists S_{R_j} = \{S_{R_j}^1, S_{R_j}^2, \dots, S_{R_j}^{N_{R_j}}\}, j = \overline{1, N_R}. \quad (2)$$

Переход НИОКР и ресурсов в различные состояния происходит под действием внешних и внутренних событий, определяемых задачами стратегии развития и научно-техническим потенциалом организации.

Состояние программы НИОКР в любой момент времени $t \in T$ (T – типовой интервал деятельности организации: год, квартал) представляет собой точку в пространстве

$$S_{\Pi} \subseteq S_{P_1} \times S_{P_2} \times \dots \times S_{P_{N_P}}, \quad (3)$$

где $\Pi = \{P_1, P_2, \dots, P_{N_P}\}_t$, $t \in T$ – множество работ, выполняемых организацией на момент времени t .

Аналогичное выражение справедливо для ресурсов, т.е.

$$S_{\mathcal{R}} \subseteq S_{R_1} \times S_{R_2} \times \dots \times S_{R_{N_R}}, \quad (4)$$

где $\mathcal{R} = \{R_1, R_2, \dots, R_{N_R}\}_t$, $t \in T$.

Формальные описания состояний (3) и (4) позволяют представить процесс научно – производственной деятельности организации в виде

$$F_{TEK} = \langle S_{TEK}, \Pi \rangle, \quad (5)$$

где $S_{TEK} \subset S_{\Pi} \times S_{\mathcal{R}}$ – выражение, определяющее текущее распределение ресурсов по работам.

Рассмотрим возможность введенного формального аппарата для описания характерных состояний работ в процессе их реализации.

К факторам, существенно влияющим на процесс успешного проведения НИОКР, относятся:

- критические участки работы – такое ее состояние $S_{P_i}^j \in S_{P_i}$, которое предполагает использование неделимego ресурса R_j ;
- блокирование работы в состоянии $S_{P_i}^j \in S_{P_i}$, если для текущего состояния организации не существует ни одного состояния $S_{P_i}^q \in S_{P_i}$, позволяющего продолжить (завершить) данную работу $S_{P_i}^j \rightarrow S_{P_i}^q$. конфликт – пребывание двух или более работ $P_i, P_r, \dots$ в заблокированных состояниях $S_{P_i}^j \in S_{P_i}$, $S_{P_r}^j \in S_{P_r}$ по причине их вхождения в критические участки относительно ресурса R_j ;
- тупиковое состояние заблокированной работы $S_{P_i}^j \in S_{P_i}$, если она заблокирована во всех состояниях независимо от текущей реализации программы НИОКР $S_{TEK}^r \in \Pi_i$.

Заблокированная работа P_i находится в тупиковом состоянии $S_{P_i}^j \in S_{P_i}$, если она заблокирована во всех состояниях независимо от текущей реализации программы НИОКР $S_{TEK}^r \in \Pi_i$.

Возникновение в процессе деятельности организации конфликтных и тупиковых ситуаций приводит к срыву сроков, снижению качества выполнения НИОКР и, как следствие, снижению эффективности деятельности организации. Поэтому управление организацией должно строиться на основе оптимальной стратегии управления НИОКР.

Ситуационные стратегии управления НИОКР

Эффективность выполнения НИОКР достигается при использовании следующих стратегий:

- исключение возможности возникновения тупиков в ходе выполнения работ для типового интервала деятельности организации на этапе планирования распределения ресурсов;
- минимизация числа тупиковых ситуаций, их своевременное обнаружение и восстановление нормального ритма выполнения работ.

Причем необходимо выбрать такую стратегию управления ограниченными ресурсами при выполнении НИОКР, при которой на типовом интервале достигается наибольшая эффективность научно-производственной деятельности организации.

В общем случае первая стратегия управления НИОКР, принципиально исключающая возникновение тупиковых ситуаций, как правило, не позволяет выполнить работы в установленные сроки, не обеспечивает высокой степени использования ресурсов организаций и требует значительных затрат на реализацию [1]. Кроме того, стохастический характер априорной информации о выполнении работ не позволяет использовать стратегии управления в реальном времени. Поэтому выбор наилучшей стратегии управления НИОКР связан, в основном, с минимизацией числа кон-

фликтных ситуаций, разработкой эффективных алгоритмов распознавания тупиковых ситуаций и выхода из них.

Известные алгоритмы выявления тупиковых ситуаций характеризуются или высокой ресурсоёмкостью, или ограниченностью по своим функциональным возможностям. Предлагаемый алгоритм эффективен для наиболее общего случая распределения повторно используемых разделяемых ресурсов.

Если представить текущее состояние выполнения НИОКР взвешенным графом, то распределение ресурсов можно описать соотношением

$$\forall j \left(\left(\sum_{i=1}^{N_P} \bar{\lambda}_{ji} + \sum_{i=1}^{N_P} \lambda_{ij} \right) \leq w_j \right), \quad (6)$$

где $\bar{\lambda}_{ji}$ - элемент ресурса j -го типа, выделенный работе P_i ; λ_{ij} - элемент ресурса j -го типа, запрашиваемый работой P_i .

При этом любая работа P_i приостанавливается, если она выдала запрос на ресурс R_i , причём

$$\sum_{i=1}^{N_P} (\bar{\lambda}_{ji} + \lambda_{ij}) > w_j. \quad (7)$$

Текущее состояние работ и ресурсов полностью характеризуется матрицей распределения ресурсов

$$W_P = \|w_{ji}^P\|, \quad i = \overline{1, N_P}, \quad j = \overline{1, N_R}, \quad (8)$$

где w_{ji}^P - количество ресурсов j -го типа, выделенных работе P_i и

$$w_{ji}^P = \sum_{(R_j, P_i)} \bar{\lambda}_{ji}$$

(сумма берётся по всем дугам, соединяющим вершины R_j и P_i , как показано на рис. 2); матрица запросов на ресурсы записывается

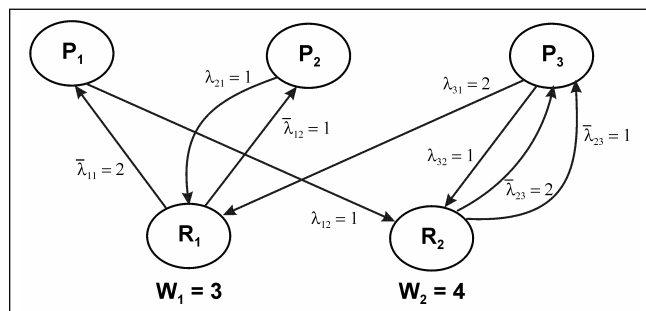
$$W_3 = \|w_{ij}^3\|, \quad i = \overline{1, N_P}, \quad j = \overline{1, N_R}, \quad (9)$$

где w_{ij}^3 - количество ресурсов j -го типа, запрашиваемых работой P_i и

$$w_{ji}^3 = \sum_{(R_j, P_i)} \lambda_{ij};$$

вектор наличных ресурсов организации составляет

$$W = (w_1, w_2, \dots, w_{N_R}); \quad (10)$$



$$W = (3, 2), \quad W_{CB} = (0, 1).$$

Рис. 2. Представление работ и ресурсов взвешенным графом

вектор свободных ресурсов определяется выражением

$$W_{CB} = (w_{CB}^1, w_{CB}^2, \dots, w_{CB}^{N_R}), \quad (11)$$

$$\text{где } w_j^{CB} = w_j - \sum_{i=1}^{N_P} w_{ji}^P.$$

Очевидно, что с введением W_P, W и W_{CB} по соотношениям (8) – (11) условие (7) сводится к виду $\lambda_{ij} > w_j^{CB}$.

Если некоторая работа P_i заблокирована по ресурсу R_j в соответствии с условием (7), то будем считать, что она блокируется всякой незаблокированной работой P_q , для которой выполняется условие $w_{jq}^P \neq 0$. В соответствии с этим условием строится матрица блокировки работ

$$B = \|b_{kl}\|, \quad k, l = \overline{1, N_P}, \quad (12)$$

где

$$b_{kl} = \begin{cases} 1, & \text{если работа } P_k \text{ заблокирована работой } P_l \\ & \text{по некоторому ресурсу } R_j, \text{ т.е. } \lambda_{kj} > w_j^{CB} \wedge w_{jl}^P \neq 0; \\ 0 & \text{– в противоположном случае.} \end{cases}$$

Алгоритм поиска тупиковых ситуаций основан на последовательном применении двух итерационных процедур сокращения и коррекции матрицы блокировок B . Если матрица блокировок работ не является полностью сокращаемой, то выполнение работ в организации находится в тупиковом состоянии.

Практическая реализация и внедрение результатов исследований

Предложенная методика реализуется в НИР по созданию распределенной интеллектуальной системы поддержки принятия решений по управлению процессами создания техники в ходе проводимых в институте НИОКР.

Предложенная стратегия управления использована в САПР института ЗАО «МНИТИ» и обеспечивает оперативное обнаружение и исключение тупиковых ситуаций в автоматизированном режиме.

Литература

1. Назаров С.В. Операционные системы специализированных вычислительных комплексов: Теория построения и системного проектирования. – М.: Машиностроение, 1989. – 400 с.
2. Ануфриев И.К., Вилкова Н.Н. Автоматизация экономического управления предприятием в условиях налоговой системы // Научно – технический сборник «Техника средств связи». – М.: 1991.
3. Ануфриев И.К., Бурков В.Н., Вилкова Н.Н., Рапацкая С.Т. Модели и механизмы внутрифирменного управления. – Препринт. – ИПУ РАН. – 1994.
4. Барсуков А.Г., Дудко В.К. Направления автоматизации процессов макроуправления созданием образцов техники//Системы и средства связи, телевидения и радиовещания. – 2005. - №1,2. – С. 29 – 32.
5. Г.Я. Гольдштейн. Стратегические аспекты управления НИОКР. Таганрог: ТГРУ, 2000. – 244 с.

МОДЕРНИЗАЦИЯ РЛС НА ОСНОВЕ СОВРЕМЕННЫХ ТЕХНОЛОГИЙ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ

А.А. Кузин, А.В. Миронов, А.Г. Рындык, С.Б. Сидоров, В.Д. Ястребов

Введение

Развитие технологий цифровой обработки сигналов (ЦОС) существенным образом изменило подход разработчиков к проектированию систем ЦОС радиолокационных систем. В настоящее время такие системы выполняются в виде аппаратно-программных комплексов, включающих управляющий компьютер, систему внутри и межпериодной цифровой обработки сигналов, а также рабочие места операторов и устройства сопряжения с внешними системами. Система внутри и межпериодной ЦОС строится на основе использования быстродействующих аналого-цифровых преобразователей, программируемых логических интегральных микросхем и цифровых сигнальных процессоров.

В данной работе приведен опыт практической разработки такой системы коллективом сотрудников с участием авторов статьи. Работа выполнялась в рамках работ по модернизации действующего парка РЛС. Требовалось реализовать обработку сигналов в двухчастотной трехкоординатной РЛС кругового обзора. Для измерения угла места использовался фазовый метод измерения, что потребовало реализовать два канала приема на каждой частоте. Кроме того, системы защиты должны были обеспечивать традиционную для обзорных РЛС защиту от активных шумовых (АШП) импульсных и пассивных помех. Для защиты от АШП использовались две компенсационные антенны. Всего система содержала восемь основных и восемь компенсационных каналов приема.

Общая Архитектура системы

При выборе архитектуры системы авторы руководствовались необходимостью разработать аппаратно-программный комплекс, который может быть легко адаптирован для широкого класса радиолокационных применений. Разработанную систему (рис.1) можно классифицировать как централизованную архитектуру, управляемую от центрального компьютера, на котором реализуется вторичная обработка радиолокационной информации (РЛИ), взаимодействие с рабочими местами операторов, получение данных и управление режимами функционирования блока цифровой обработки сигналов, управление и контроль радиолокатора, а также сопряжение с внешними системами.

Обобщен практический опыт разработки системы цифровой многоканальной обработки радиолокационных сигналов. Система выполнена в виде аппаратно-программного комплекса, который может быть легко адаптирован для широкого класса радиолокационных применений и имеет централизованную архитектуру, управляемую от центрального компьютера, на котором реализуется вторичная обработка радиолокационной информации и управление режимами функционирования блока ЦОС, а также сопряжение с внешними системами. Блок внутри и межпериодной ЦОС выполнен в стандарте compact PCI на FPGA и DSP процессорах и обеспечивает аналого-цифровое преобразование, согласованную фильтрацию, защиту от импульсных, активных шумовых и пассивных помех, обнаружение и вычисление параметров принимаемых сигналов в четырех когерентных каналах обработки.

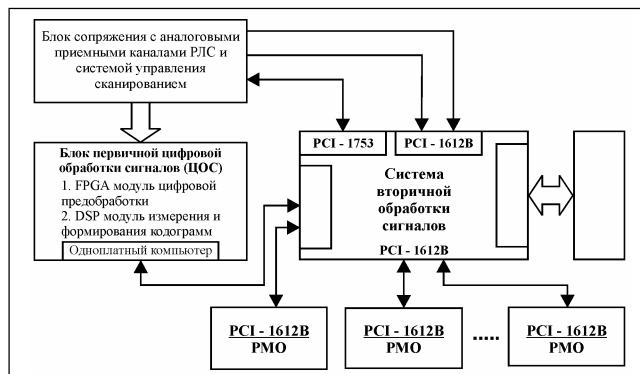


Рис.1 Архитектура системы обработки

В качестве линий связи использовался последовательный интерфейс RS-422 для обмена данными и RS-485 для обмена с датчиками системы сканирования. В качестве компьютеров вычислительной системы и рабочих мест операторов использовались промышленные компьютеры класса Pentium 4 под управлением операционной системы Linux. При необходимости они могут быть заменены на любые аналогичные, обладающие требуемой производительностью и имеющие в своём составе адаптеры последовательных портов для сопряжения между собой.

Блок ЦОС был реализован в крейте Compact PCI и включал: одноплатный компьютер (ОК) с электронным флеш-диском под управлением ОС Linux; четыре платы цифровой обработки на FPGA-процессорах (использовались ПЛИС семейства Stratix фирмы Altera); две платы цифровой обработки на DSP процессорах L1879BM1 (использовались платы МЦ 4.04 производства НТЦ «Модуль»). Выбор такой архитектуры позволил получить высокую производительность коммуникаций за счёт использования шины PCI для обмена между ОК и МЦ 4.04

и дополнительной шины для объединения между собой плат цифровых приемников и DSP модулей МЦ 4.04. Кроме того, выбранный конструктив имеет высокую механическую и электрическую надёжность и имеет аналогии в разрабатываемой отечественной аппаратуре для жёстких условий эксплуатации. Использование трех различных типов вычислительных элементов позволяет сочетать высокую производительность FPGA-процессоров, хорошую программируемость сигнальных процессоров с возможностями отладки и лёгкостью сопряжения одноплатного управляющего компьютера крейта.

Вся обработка была условно разделена на три этапа. Первый включал быстроедействующее аналого-цифровое преобразование на промежуточной частоте, цифровую обработку принимаемых сигналов в соответствии с алгоритмами обработки (цифровую согласованную фильтрацию, защиту от импульсных, активных шумовых и пассивных помех, амплитудное детектирование и межпериодное накопление). Эта обработка выполнялась на специально разработанных четырехканальных платах цифрового приема и заканчивалась формированием кодограмм, содержащих отсчеты модулей и фаз комплексных огибающих сигналов во всех каналах обработки.

На втором этапе производилось нормирование вероятности ложных тревог, формирование кодограмм сигналов, превысивших порог обнаружения, и по этим кодограммам производилось вычисление параметров принимаемых сигналов, включая дальность, азимут и угол места цели. Для измерения угла места использовался

фазовый метод измерения, в соответствии с которым угол места измерялся по фазовому сдвигу сигналов в двух каналах обработки, имеющих смещенные фазовые центры в вертикальной плоскости.

Для реализации второго этапа обработки были использованы 4-х процессорные платы МЦ4.04 на процессорах Л1879ВМ1. В результате второго этапа обработки формировались кодограммы обнаруженных отметок, которые являлись исходными для третьего этапа (траекторной обработки), реализованного на компьютере класса Pentium 4.

Модуль ЦОС на FPGA-процессорах

Модули цифрового приема на FPGA процессорах были изготовлены ЗАО «Инструментальные системы». На каждой плате было реализовано 4 когерентных канала приема для обработки сигналов двух частотных и двух фазовых приемных каналов. Аналогичная плата использовалась для обработки сигналов 4-х компенсационных каналов (по два компенсационных канала на каждой частоте). Таким образом, для обработки сигналов одной антенны (во всем комплексе их могло быть две) требовалось две платы на FPGA-процессорах.

На рис.2 показана структура обработки, реализованной на платах с FPGA-процессорами. В цифровом приемнике после дискретизации на промежуточной частоте (30МГц) осуществляется формирование отсчетов комплексной огибающей принимаемого сигнала и понижение частоты дискретизации до значения, определяемого шириной спектра сигнала (300кГц).

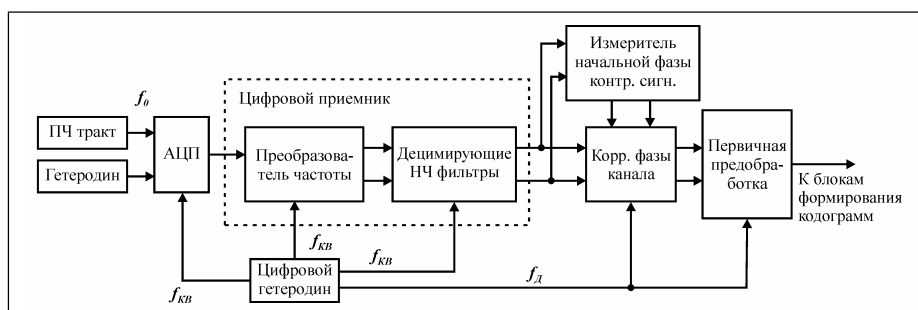


Рис. 2. Структура канала цифрового приема многоканальной РЛС.

В каждом из основных каналов после аналого-цифрового преобразования и амплитудно-фазовой коррекции (АФК) каналов с использованием контрольного сигнала осуществляется: бланкирование боковых лепестков диаграммы направленности антенны (ПСБЛ); адаптивная двухканальная компенсация шумовых помех

(АКШП) с бланкированием несинхронных импульсных помех (НИП); адаптивная череспериодная компенсация пассивных помех, амплитудное детектирование (АД) и некогерентное накопление (НН) сигнальных пакетов.

Структура системы защиты от комбинированных помех в одном частотном канале приведена на рис.3.

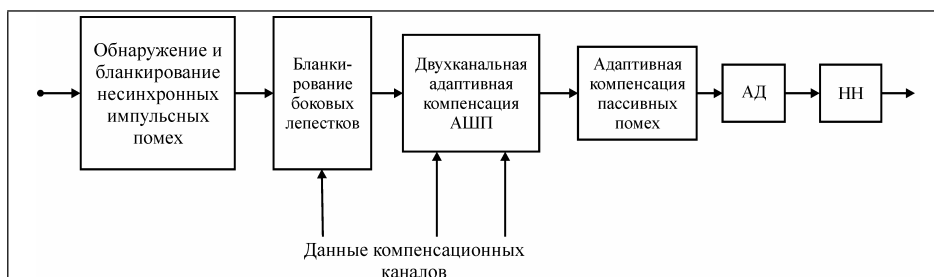
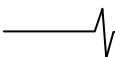


Рис.3. Структура системы защиты от комбинированных помех.



Кроме рассмотренной обработки на платах с FPGA-процессорами была реализована межканальная ЦОС, включающая: вычисление взаимных корреляционных моментов комплексных огибающих в фазовых каналах, соответствующих двум частотам приема и несущим информацию об угле места цели; межпериодного корреляционного момента комплексных огибающих в каждом частотном канале, несущего информацию о частоте Доплера цели; взаимных корреляционных моментов, несущих информацию об угле места источника АШП и оценок дисперсии АШП и межканальных корреляционных моментов АШП, необходимых для работы пеленгационного канала.

В результате обработки на платах с FPGA-процессорами формировались данные всех каналов дальности с частотой дискретизации, определяемой шириной спектра сигнала. Эти данные передавались для последующей обработки на платы с DSP процессорами.

Процессорный модуль на DSP процессорах

На платах с DSP процессорами (МЦ 4.04) были реализованы алгоритмы нормирования вероятности ложных тревог и измерения параметров обнаруженных целей. В результате обработки формировались кодограммы для каждой обнаруженной цели, содержащие следующие параметры: время задержки сигнала (дальность до цели); азимут цели; угол места цели (угломестный фазовый сдвиг); амплитуду сигнала; длину сигнального пакета; доплеровский набег фазы за период зондирования.

Обработка на DSP процессорах заканчивались передачей кодограмм на систему вторичной обработки РЛИ.

Модуль вторичной обработки РЛИ

Модуль вторичной обработки предназначен для решения задач вторичной обработки радиолокационной информации (РЛИ) в реальном времени. Основными из них являются следующие задачи:

- траекторная обработка РЛИ о воздушных объектах;
- определение государственной принадлежности;
- распознавание воздушных объектов;
- регистрация РЛИ, результатов вторичной обработки и действий операторов на рабочих местах;
- сопряжение с внестанционными потребителями вторичной информации;
- взаимодействие с другими модулями системы (рабочими местами операторов, модулем ЦОС);
- контроль состояния системы.

Конструктивно модуль «Обработка» был выполнен на промышленном компьютере класса Pentium 4. Взаимодействие с модулем ЦОС осуществляется по последовательному интерфейсу RS-422 с использованием платы последовательного интерфейса PCI-1612B. Взаимодействие с набором рабочих мест операторов осуществляется по последовательным интерфейсам RS-422 и RS-485 (для выносных РМО) с использованием двух плат последовательного интерфейса PCI-1612B. Взаимодействие с блоком сопряжения с антен-

ными устройствами осуществляется по последовательному интерфейсу RS-485 с использованием платы последовательного интерфейса PCI-1612B, а также платы PCI-1753. Сопряжение с внестанционными потребителями осуществляется посредством программируемого канального модуля ПКМ386-4 с шиной ISA.

Модуль «Обработка» может работать как в штатном режиме, так и автономном, при отсутствии какой-либо аппаратной связи с другими компонентами системы. В автономном режиме РЛИ формируется подсистемой имитации, входящей в состав модуля. Кроме этого предусмотрена возможность имитации одного РМО с помощью специальной встроенной подсистемы. При этом моделируется поступление с РМО различных команд управления.

При выборе архитектуры программного обеспечения модуля «Обработка» принимались в расчет следующие факторы:

- значительное число решаемых модулем задач с различными степенями критичности выполнения;
- большое количество потребителей и источников информации с ее асинхронным поступлением;
- требование функционирования в реальном времени.

С учетом перечисленного выше программное обеспечение модуля «Обработка» представлено в виде многопоточного приложения (рис. 4).

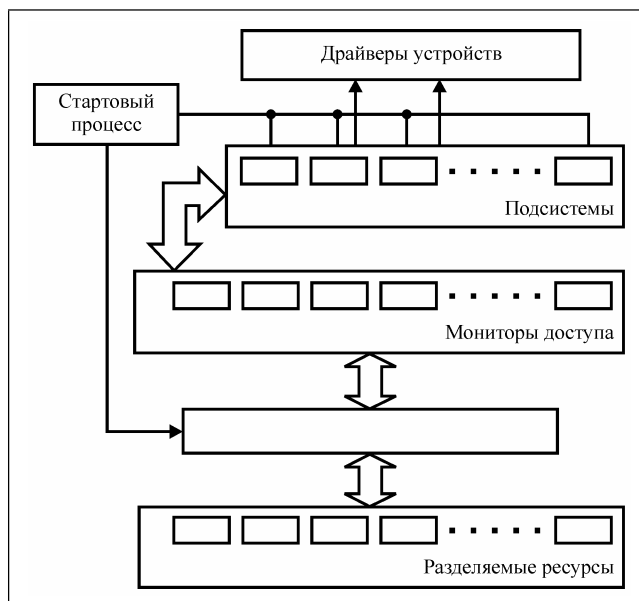


Рис.4. Архитектура ПО модуля «Обработка»

При разработке таких приложений ключевым этапом является распределение задач между потоками и организация их взаимодействия. В основу ПО модуля «Обработка» положена одна из основных моделей подобного рода приложений – модель «без лидера». Эта модель находит свое применение, когда набор источников поступления информации строго определен, и управляющая подсистема отсутствует. В рамках данной модели каждая подсистема отвечает за свой ввод (обнаружение наступления события), обеспечивает требуемую обработку события и инициирует наступление событий для других подсистем приложения.

Программное обеспечение модуля «Обработка» выполнено под управлением операционной системы Linux, с ядром версии 2.4.25. При этом каждая подсистема представлена отдельным потоком управления thread («облегченным процессом»). Важнейшим из требований, предъявляемых к системам реального времени, является быстрая реакция на наступление события. В тоже время общеизвестно, что операционная система Linux не является системой реального времени, поскольку обычные пользовательские процессы могут выполняться без вытеснения ядром до 10 миллисекунд в силу принятой в операционной системе схемы диспетчеризации. Вместе с тем наряду с «обычными» процессами могут быть созданы специальные процессы с ненулевыми статическими приоритетами. Эти приоритеты назначаются в соответствии с требованиями, предъявляемыми к конкретной задаче. Планировщик операционной системы Linux осуществляет планирование таких процессов другими дисциплинами, в частности дисциплиной «циклическое RR-планирование». Данная дисциплина применяется для процессов реального времени и обеспечивает более быстрый отклик высокоприоритетных процессов по сравнению с обычными процессами. В ПО модуля «Обработка» для потоков управления была использована рассмотренная дисциплина планирования, которая для задачи модуля обеспечивает приемлемые времена отклика на события.

Взаимодействие подсистем по управлению и обменом данными осуществляется через «разделяемые ресурсы», представленные общими для потоков областями памяти. Каждый ресурс имеет свой идентификатор, структуру и предназначен для хранения определенной информации. Для доступа к ресурсам подсистемы используют мониторы Хоара, по одному на каждый вид ресурса. Эти интерфейсные элементы предоставляют безопасный и надежный доступ к операциям над ресурсами через свои методы. Специальный программный контроллер обеспечивает централизованное управление всеми ресурсами и операциями над ними. При этом контроллер позволяет одновременный доступ к нему через мониторы Хоара со стороны нескольких подсистем.

Модуль РМО

Модуль «Рабочее место оператора» предназначен для решения задач отображения различной радиолокационной информации и управления функционированием системы со стороны оператора в реальном времени. Основными из них являются следующие задачи:

- отображение первичной и вторичной РЛИ о воздушных объектах (ВО);
- поддержка режимов отображения – масштабирование, смещенный центр, центр по ВО;
- взаимодействие с модулем «Обработка»;
- управление работой модуля ЦОС;
- управление встроенным имитатором радиолокационной обстановки модуля «Обработка»;
- управление подсистемой вторичной обработки модуля «Обработка»;

- отображение диагностической информации по состоянию системы.

Конструктивно каждый модуль «РМО» реализован на ПЭВМ Pentium 4. Взаимодействие с модулем «Обработка» осуществляется по последовательному интерфейсу RS-422 для локальных РМО и RS-485 для выносных РМО с использованием платы последовательного интерфейса PCI-1612B (рис.1).

Модуль «РМО» работает в режиме взаимодействия с модулем «Обработка». Программное обеспечение модуля «РМО» представлено комплексом взаимодействующих процессов (рис.5). В модулях «РМО» и «Обработка» задействован широкий спектр один и тех же компонент, таких как математические преобразования и кодограммы обмена. Это позволило с одной стороны существенно сократить общее время разработки, а с другой упростить внесение изменений в систему. Программное обеспечение модуля «РМО» выполняется под управлением операционной системы Linux, с ядром версии 2.4.25. В качестве графической среды используется X Window System и библиотека OpenMotif.

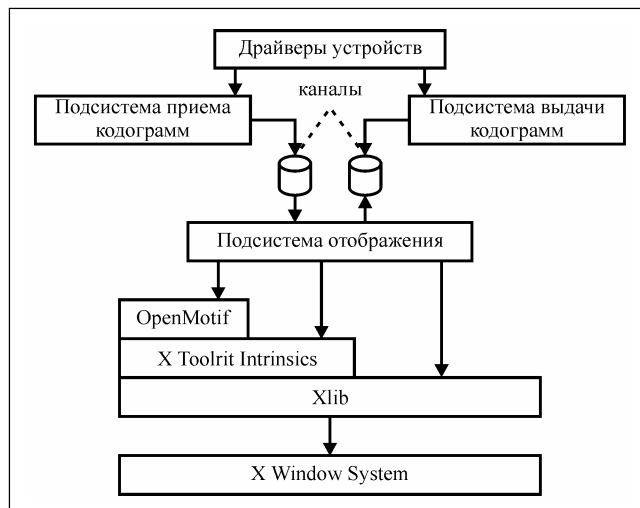


Рис. 5. Архитектура ПО модуля «РМО».

Заключение

Опыт выполненной разработки показал, что на основе современных технологий ЦОС возможно создание эффективной и относительно недорогой системы цифровой обработки и отображения радиолокационной информации обзорной РЛС.

Использование ПЛИС семейства Stratix фирмы Altera позволило реализовать на одной плате 4-х каналный цифровой приемник со всем комплексом традиционных систем защиты от активных шумовых, импульсных и пассивных помех.

Система стабилизации вероятности ложных тревог (СУЛТ) и измерение параметров целей были реализованы на DSP процессорах Л1879ВМ1. Их производительности в составе 4-х процессорной платы МЦ4.04 оказалось достаточно для сопряжения цифровых приемников с системой вторичной обработки и обеспечения траекторной обработки и отображения информации на рабочих местах комплекса в реальном времени.

ВЫЧИСЛИТЕЛЬНЫЙ МАКРОКОНВЕЙЕР С ПЕРЕМЕННЫМ ТАКТОМ РАБОТЫ

Г.В. Зайцев

Введение

Конвейеризация является одним из основных способов повышения производительности вычислительных систем и широко используется на всех уровнях структурной иерархии устройств. В данной работе рассматривается верхний уровень указанной иерархии, а конвейерная структура этого уровня называется макроконвейером. Анализируемая система (рис. 1) содержит ряд последовательно соединенных процессоров (Пр), разделенных буферными запоминающими устройствами. Каждый из процессоров является вычислительным устройством, содержащим при необходимости внутреннее рабочее запоминающее устройство (ЗУ). На вход макроконвейера поступают последовательно во времени блоки

Рассматривается вычислительная система с конвейеризацией на уровне задач, содержащая несколько последовательно соединенных процессоров с буферными запоминающими устройствами (БЗУ) между ними. Предлагаются алгоритмы управления конвейером и рассчитывается необходимый объем БЗУ. Анализируются два способа построения БЗУ: в виде двухпортовой памяти и в виде памяти из нескольких банков. Рассматриваются временные диаграммы прохождения задач по конвейеру.

данных, относящиеся к последовательно решаемым независимым задачам. Над каждым блоком данных в каждом процессоре производятся необходимые операции обработки. В принципе, БЗУ могут отсутствовать, однако, если процессоры полностью используют как объем, так и быстродействие внутренних ЗУ, то отсутствие БЗУ приводит к резкому снижению пропускной способности макроконвейера. Поэтому в данной работе рассматривается конфигурация макроконвейера, содержащего БЗУ.

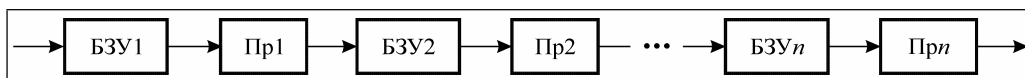


Рис. 1. Структурная схема макроконвейера

Организация БЗУ не вызывает затруднений в случае, когда макроконвейер решает одну и ту же повторяющуюся задачу над последовательно поступающими блоками данных. При этом обычно каждое БЗУ строится в виде ЗУ с двумя банками, один из которых принимает данные от предыдущего процессора, а второй выдает данные на последующий процессор. Для следующего блока данных (следующей задачи) банки меняются местами.

Однако в ряде случаев возникает необходимость обработки в макроконвейере различных задач, требующих различного времени решения в процессорах. Организация БЗУ при этом несколько усложняется и является предметом рассмотрения настоящей работы.

Приведем один из примеров подобного макрокон-

вейера, относящийся к области цифровой обработки информации в радиолокационных станциях (РЛС). Возможная структура устройства обработки принимаемых сигналов приведена на рис. 2. На вход аналого-цифрового преобразователя (АЦП) поступают прошедшие необходимую аналоговую обработку сигналы. Информация от одного зондирования накапливается в БЗУ1 и далее передается в программируемый процессор сигналов (ППС), выполняющий его первичную обработку (оптимальную фильтрацию сигнала и выделение отметок, превысивших порог). Результаты передаются через БЗУ2 в процессор вторичной обработки (ПВО), реализующий алгоритмы работы с обнаруженными отметками (обнаружение объектов, завязка трасс, сопровождение).



Рис. 2. Макроконвейер обработки сигнала в РЛС

Требуемая производительность и пропускная способность указанных процессоров существенно различны, также, как и характер выполняемых алгоритмов. В связи с этим они, как правило, реализуются в различных устройствах, образующих макроконвейер рис. 2.

Поток сигналов на вход такого устройства определя-

ется сменой режимов зондирования РЛС и в современных многофункциональных РЛС является довольно неравномерным. На рис. 3 приведен типичный фрагмент временной диаграммы работы импульсной РЛС с фазированной антенной решеткой. На верхней диаграмме показаны излучаемые передатчиком импульсы, на ниж-

ней – интервалы приема (набора информации в БЗУ1). Перед каждым тактом зондирования луч антенны устанавливается в необходимое угловое положение. В такте 1 зондирования производится обзор пространства с помощью импульса с линейной частотной модуляцией.

Период такого обзорного зондирования определяется максимальной дальностью РЛС и, как правило, довольно значителен, а интервал набора информации в БЗУ1 охватывает всю просматриваемую дальность.

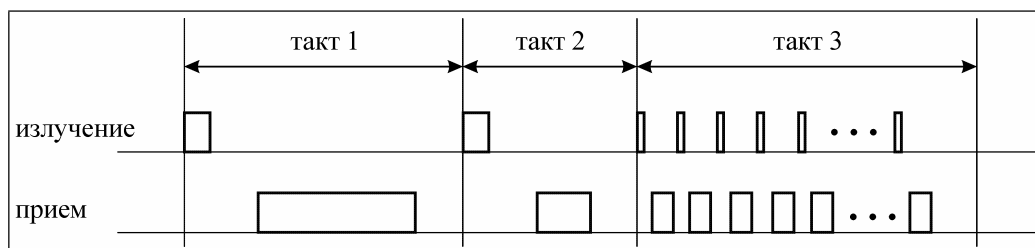


Рис. 3. Временная диаграмма работы РЛС

Такт 2 зондирования используется для сопровождения ранее обнаруженного объекта с помощью импульса с фазокодовой манипуляцией. При небольшой дальности до объекта длительность этого такта существенно меньше, чем такта 1, а интервал набора информации, в силу известной дальности, также невелик.

Такт 3 используется для сопровождения низко летящего объекта с помощью квазинепрерывного сигнала с высокой частотой повторения, позволяющего при большой длительности такта эффективно подавить отражения от местных предметов. При этом внутри такта 3 присутствует целый ряд интервалов набора информации в соответствии с числом излучаемых импульсов. Таким образом, на вход макроконвейера приемного устройства через неодинаковые интервалы времени поступают массивы информации различного объема, требующие различной обработки.

Очевидный и часто используемый способ организации такого конвейера состоит в требовании к процессорам обеспечить решение всех задач за длительность минимального такта зондирования. Однако это может более, чем на порядок, увеличить необходимую производительность процессоров, которые к тому же большую часть времени будут простаивать. В данной работе предлагается более экономный способ организации описанного макроконвейера.

Приведенные ниже методы организации макроконвейера используются с середины 80-х годов в текущих разработках процессоров радиолокационных сигналов.

Базовая модель макроконвейера

Рассмотрим модель описанного выше макроконвейера, являющуюся основой последующего анализа. Для сокращения текста в дальнейшем изложении термин конвейер используется в качестве синонима термина макроконвейер. Назовем этапом конвейера БЗУ с последующим за ним процессором, а временной интервал, за который обрабатываются данные некоторой задачи в любом устройстве конвейера, – соответствующим тактом работы этого устройства. Более точно, тактом работы процессора, относящимся к некоторой задаче, назовем временной интервал, за который данные этой задачи обрабатываются в рассматриваемом процессоре, включая время приема входных данных и выдачи результа-

тов. Тактом работы БЗУ по записи (соответственно, считыванию), относящимся к решению некоторой задачи, назовем временной интервал, в течение которого может производиться запись (считывание) данных этой задачи. Очевидно, такты работы любого БЗУ по записи (считыванию) должны согласовываться с тактами работы предыдущего (последующего) процессоров конвейера.

В частности, в каждый такт работы входного устройства (входной такт) на вход конвейера поступают входные данные очередной задачи. При этом различные задачи будем считать независимыми и никак не связанными между собой. Например, для конвейера приемного устройства такты, показанные на рис. 3, являются входными тактами.

Рассматриваемая модель конвейера описывается следующими условиями.

У1. Длительность входных тактов конвейера может принимать любое значение от нуля до T_m .

У2. Длительность тактов обработки информации фиксированной задачи одинакова для всех процессоров конвейера и равна длительности входного такта этой задачи.

У3. Максимально возможный объем V_i информации, поступающей на вход i -го этапа конвейера и относящейся к фиксированной задаче, пропорционален длительности T входного такта этой задачи, т. е.

$$V_i = k_i T, \quad i = 1, 2, \dots, n,$$

где k_i – константы, а n – количество этапов конвейера.

У4. Обработка данных фиксированной задачи в процессоре некоторого этапа конвейера начинается только после окончания ее обработки на предыдущем этапе и после окончания обработки предыдущей задачи в рассматриваемом процессоре.

Условие У1 говорит о том, что произвольно длинные входные такты невозможны. В противном случае объем входного БЗУ стал бы бесконечным, что не имеет смысла. Условие У2 является основой конвейерного режима работы, обычно для конвейерных систем и обеспечивает минимально возможную производительность процессоров. Условие У3 о пропорциональности объема входной информации длительности такта является хорошим приближением для ряда систем, например, для описанного выше конвейера приемного устройства. Очевидно, условием 3 могут быть описаны любые конвейеры при соот-

ветствующем выборе констант k_i .

Условие У4 продиктовано стремлением к упрощению управляющих устройств процессоров. Отказ от условия У4 может привести к конвейерному способу обработки задачи сразу на нескольких процессорах. Этот случай может быть сведен к рассматриваемому путем объединения таких процессоров в один укрупненный процессор.

В последующих разделах обсуждается установившийся режим работы конвейера, описываемый следующими условиями.

У5. На входе конвейера такты следуют друг за другом без интервалов между ними.

У6. До рассматриваемого интервала времени закончена обработка хотя бы одной задачи с длительностью входного такта T_m .

Эти условия в сочетании с условиями У2, У3 приводят к полной и постоянной загрузке процессоров конвейера. Кроме того, следствием условия У6 является тот факт, что задержка информации любой задачи на каждом этапе конвейера не может быть меньше T_m . Действительно, при решении задачи с длительностью входного такта T_m , эта задержка, по условию У2, имеет величину не менее T_m , а в силу полной и постоянной загрузки процессоров не может сократиться и для всех последующих задач.

С другой стороны, очевидно, что задержки величиной T_m достаточно для реализации конвейерного режима. Увеличение этой задержки сверх минимально необходимой приводит к увеличению необходимого объема БЗУ. Поэтому дальнейшее рассмотрение установившегося режима проводится при следующем предположении.

У7. В установившемся режиме задержка информации на каждом этапе конвейера равна минимально возможной величине T_m .

Рассмотрим теперь способы управления конвейером и необходимый объем БЗУ. При этом ограничимся только такими способами построения БЗУ, при которых БЗУ всегда доступно предыдущему процессору на запись, а следующему – на считывание. Это условие приводит к отсутствию конфликтов при обращении к БЗУ и максимизирует пропускную способность конвейера.

Проанализируем сначала установившийся режим работы, так как именно он предъявляет максимальные требования к объему БЗУ.

Двухпортовое БЗУ

Первый очевидный способ организации БЗУ состоит в использовании двухпортового ЗУ с независимыми обращениями по входу (запись) и выходу (считывание). При этом алгоритм работы БЗУ элементарен и состоит в следующем.

Алгоритм 1. Запись. Записываемая информация последовательно и циклически заполняет имеющийся объем БЗУ.

Считывание А. Считывание в процессор из БЗУ информации очередной задачи начинается немедленно после выполнения условия У4: обработка данной задачи закончена в предыдущем процессоре и обработка предыдущей задачи закончена в рассматриваемом процессоре.

Считывание С. Такт считывания из БЗУ очередной задачи начинается через время T_m после начала записи этой задачи.

Для установившегося режима оба способа считывания эквивалентны из-за полной и постоянной загрузки процессоров. Разделение этих способов необходимо при рассмотрении переходных режимов. Далее будет показано, что способ считывания А (соответственно, способ С) определяет асинхронный (синхронный) режим работы конвейера, откуда и происходит мнемоническое обозначение этих способов.

Единственной задачей при использовании двухпортового БЗУ является обеспечение достаточного его объема.

Утверждение 1. При использовании двухпортового ЗУ в качестве БЗУ i -го этапа конвейера необходимый объем БЗУ составляет не более $3k_i T_m$, $i = 1, 2, \dots, n$.

Для доказательства рассмотрим временные диаграммы тактов записи и считывания i -го БЗУ, приведенные на рис. 4. Каждый временной такт обозначен прямоугольником соответствующей длительности, а его номер указан цифрой внутри прямоугольника. Подобные же обозначения используются и на последующих рисунках. В установившемся режиме эти диаграммы имеют одинаковый вид и отличаются лишь задержкой на T_m . Поэтому можно пользоваться одной временной диаграммой, например, диаграммой записи. При этом состояние по считыванию в момент t будет определяться диаграммой записи для момента $t - T_m$.

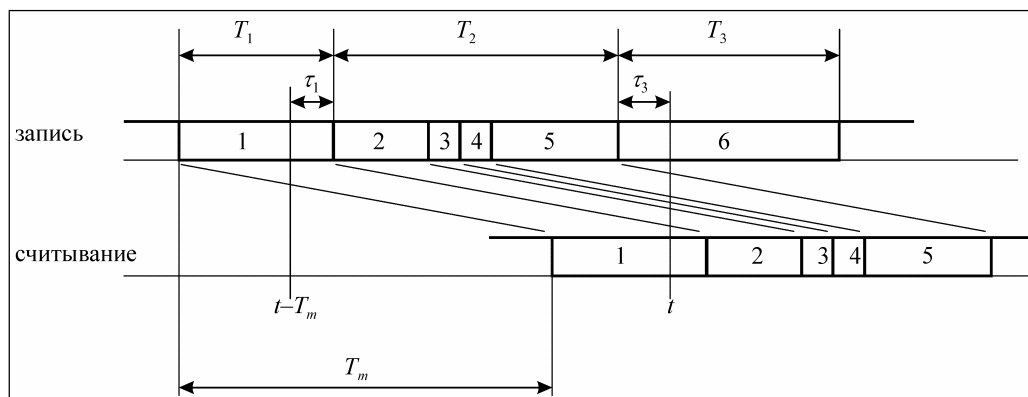


Рис. 4. Временная диаграмма макроконвейера

Определим объем информации V_t , находящийся в буфере в произвольный момент t . Он равен объему записанной к этому моменту информации W_t минус объем считанной информации R_t , причем такты, для которых считывание к рассматриваемому моменту полностью закончено, можно не учитывать. Пусть T_1 – длительность текущего такта считывания, T_2 – длительность целых тактов за интервал времени $[t - T_m, t]$, а T_3 – длительность текущего такта записи (рис. 4). Обозначим через τ_1 и τ_3 длительность отрезков тактов T_1 и T_3 , попавших внутрь интервала $[t - T_m, t]$. Тогда, обозначая через $W(x)$ и $R(x)$ соответственно объем записанной и считанной информации за интервал x , имеем

$$V_t = W_t - R_t = W(T_2) + W(T_1) + W(\tau_3) - R(T_1 - \tau_1). \quad (1)$$

Очевидно, при произвольном порядке записи и считывания $W(\tau_3) \leq W(T_3)$, а $R(T_1 - \tau_1) \geq 0$. Поэтому $V_t \leq W(T_2) + W(T_1) + W(T_3)$, и, согласно условию У2, каждое слагаемое не превосходит $k_i T_m$. Следовательно, $V(t) \leq 3k_i T_m$, что и требовалось доказать.

Из приведенного доказательства видно, что весь объем БЗУ используется лишь при чрезвычайно неблагоприятных обстоятельствах, а именно, когда интервалы T_1 , T_2 , T_3 близки к T_m . Предыдущий процессор, решающий задачу такта T_3 , почти все результаты записывает в БЗУ за интервал τ_3 , а последующий – большую часть входных данных такта T_1 считывает в конце этого такта за интервал τ_1 . Тем не менее, объем БЗУ близкий к $3k_i T_m$ в общем случае может понадобиться. Достоинством такого варианта является возможность использования процессором частей предшествующего и последующего БЗУ, выделяемых на задачу текущего такта, на протяжении всего этого такта как своей оперативной памяти.

Объем БЗУ можно существенно сократить, наложив минимальные условия на порядок записи и/или считывания. Назовем процедуру считывания из БЗУ i -го этапа конвейера равномерно ограниченной, если за интервал τ от начала такта из БЗУ считывается объем информации не менее, чем $k_i \tau$, $i = 1, 2, \dots, n$. Аналогично процедуру записи в БЗУ i -того этапа конвейера равномерно ограниченной, если за интервал τ от начала такта в БЗУ записывается объем информации не более, чем $k_i \tau$, $i = 1, 2, \dots, n$.

Утверждение 2. При равномерно ограниченных процедурах записи и считывания в БЗУ i -го этапа конвейера необходимый объем двухпортового БЗУ составляет $k_i T_m$. В случае равномерной ограниченности одной из процедур записи или считывания необходимый объем двухпортового БЗУ составляет $2k_i T_m$, $i = 1, 2, \dots, n$.

Для доказательства снова рассмотрим временную диаграмму рис. 4 и равенство (1). В силу равномерной ограниченности процедуры считывания имеем

$$R(T_1 - \tau_1) \geq k_i (T_1 - \tau_1), \quad (2)$$

а в силу равномерной ограниченности процедуры записи –

$$W(\tau_3) \leq k_i \tau_3. \quad (3)$$

Подставляя (2) и (3) в (1), получаем:

$$V_t \leq k_i (T_2 + T_1 + \tau_3 - T_1 + \tau_1) = k_i T_m.$$

Аналогично доказывается и вторая часть утверждения 2. Равномерно ограниченные процедуры записи и считывания естественным образом организуются в случае, когда внутренние ЗУ предшествующего и последующего процессоров достаточно велики, чтобы для любого такта ввести все входные данные и произвести их обработку без обращения к БЗУ. В этом случае любой такт можно начать со считывания из БЗУ всей входной информации, затем произвести полностью ее обработку и лишь затем вывести выходные результаты. При такой обработке процедуры записи и считывания, очевидно, равномерно ограничены.

В качестве примера рассмотрим конвейер обработки сигнала в РЛС (рис. 2). Входное БЗУ1, принимающее данные от АЦП, должно иметь объем не менее $2k_1 T_m$, так как процедура записи определяется внешними физическими условиями и не может быть равномерно ограничена. БЗУ2 между ПК0 и ПНО может иметь объем $k_2 T_m$ при соответствующей организации алгоритмов обработки, находящейся в руках разработчика.

БЗУ из нескольких банков однопортовых ЗУ

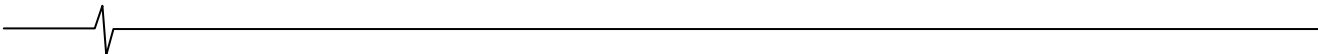
При организации БЗУ из нескольких банков однопортовых ЗУ по-прежнему необходимо обеспечить возможность независимого обращения к БЗУ по записи и по считыванию для выполнения условий работы конвейера. При произвольных процедурах записи и считывания это означает, что к некоторому банку можно обратиться по считыванию только тогда, когда запись в него закончена и производится в другой банк. Отсюда следует, что объем каждого банка БЗУ i -го этапа конвейера не имеет смысла делать больше $k_i T_m$, $i = 1, 2, \dots, n$. Действительно, при записи в один банк большего объема информации задержка между записью и считыванием превысит T_m , что противоречит условию У7.

Рассмотрим вопрос о необходимом количестве банков объемом $k_i T_m$, $i = 1, 2, \dots, n$. При этом напомним, что в данном пункте рассматривается лишь установившийся режим. Для упрощения устройства управления и удобства оперирования с банками положим, что информация любого банка должна быть записана в один банк.

Утверждение 3. Для организации БЗУ i -го этапа конвейера достаточно трех банков однопортовых ЗУ, каждый объемом $k_i T_m$, $i = 1, 2, \dots, n$.

Доказательство проведем конструктивным путем, рассмотрев алгоритм работы описанного БЗУ, не приводящий к конфликтам при обращении к нему. При этом для краткости назовем указателем записи $w(t)$ (соответственно, указателем считывания $r(t)$) целочисленную функцию, равную в момент t номеру банка, в который должно происходить текущее обращение по записи (соответственно, по считыванию).

Алгоритм 2. Запись. Поступающая информация последовательно заполняет объем каждого банка.



Указатель записи может изменяться лишь в момент смены тактов записи. При смене тактов указатель записи не изменяется, если время от начала первого такта записи в данный банк до конца начинающегося такта не превышает величины T_m ; в противном случае указатель записи получает приращение 1 (mod 3).

Способы считывания А и считывания С идентичны таковым для алгоритма 1.

Переходя к доказательству утверждения 3, заметим, что для отсутствия конфликтов в БЗУ необходимо и достаточно, чтобы в любой момент времени выполнялось неравенство $w(t) \neq r(t)$. Причем, достаточно рассмотреть лишь моменты изменения функций $w(t)$ и $r(t)$, т.е. моменты смены тактов по записи и по считыванию.

Моменту изменения $r(t)$ с i на $i+1(\text{mod } 3)$ соответствует временная диаграмма рис. 4 с $\tau_1=0$. Тогда, если $\tau_3=0$, то за интервал $(t - T_m, t)$ информация, согласно алгоритму 2, записывалась в банк $i+1(\text{mod } 3)$, полностью его заполнила и $w(t)=i+2(\text{mod } 3)$. Если же $\tau_3>0$, то информация текущего такта записи не размещается в банке $i+1(\text{mod } 3)$, и $w(t)$ переключился на $i+2(\text{mod } 3)$ в начале текущего такта записи. В любом случае в момент изменения $r(t)$ с i на $i+1(\text{mod } 3)$ выполняется неравенство $w(t) \neq r(t)$.

Моменту изменения $w(t)$ с i на $i+1(\text{mod } 3)$ соответствует временная диаграмма 4 с $\tau_3=0$. При этом за интервал $(t - T_m, t)$ не могло быть более одного изменения $w(t)$, так как с момента первого такого изменения до момента t информация может быть размещена в одном банке и по алгоритму 2 так и записывается. Тогда $r(t) = i-1(\text{mod } 3)$, если таковое переключение было внутри интервала $(t - T_m, t)$, и $r(t)$ меняет свое состояние с $i-1(\text{mod } 3)$ на $i(\text{mod } 3)$, если таковое переключение было в момент $t - T_m$. В любом случае $w(t) \neq r(t)$. Утверждение доказано.

При описанном способе организации БЗУ так же, как и для утверждения 1, части входного и выходного БЗУ для i -того процесса, относящиеся к текущему такту его работы, могут использоваться процессором на протяжении всего такта как собственная оперативная память.

Из приведенных рассуждений легко следует, что двух банков БЗУ, вообще говоря, недостаточно. В

некоторых более частных случаях за счет усложнения аппаратуры управления конвейером и накладывания довольно жестких условий на соотношения скоростей записи и считывания БЗУ можно обойтись и двумя банками. Однако в данной работе эти варианты из-за их специфичности рассматриваться не будут.

Переходные режимы работы макроконвейера

До сих пор рассматривался лишь установившийся режим работы конвейера, описываемый условиями У5–У7 и характеризующийся полной и постоянной загрузкой всех процессоров при задержке информации на каждом этапе конвейера на время T_m . К переходным режимам относятся начало и окончание работы конвейера, а также ситуации, возникающие при наличии на входе свободных от работы временных интервалов («пустых» тактов). Способы организации конвейера, рассмотренные ранее и описанные там алгоритмы и утверждения об объемах БЗУ остаются справедливыми, как нетрудно проверить, и для переходных режимов, так как установившийся режим предъявляет наибольшие требования как к производительности процессоров, так и к объему БЗУ.

Тем не менее, переходные режимы имеют свои особенности. Рассмотрим, например, начало работы конвейера. Временная диаграмма начала работы зависит от способа считывания в алгоритмах 1 и 2. При считывании типа А в случае, когда первые входные такты имеют длительность меньше T_m , задержка на каждом этапе конвейера также меньше T_m и определяется задачей наиболее длинного такта, прошедшей данный процессор (временная диаграмма рис. 5 для двухэтапного конвейера). При появлении более длинного такта задержка на каждом этапе возрастает, а в работе процессора появляются свободные интервалы (рис. 5). При считывании типа С задержка на каждом этапе фиксирована и сразу входит в установившийся режим (рис. 6). Назовем поэтому конвейер со считыванием типа А во всех БЗУ асинхронным, а со считыванием типа С – синхронным конвейером. Таким образом, для синхронного конвейера временные диаграммы тактов на всех этапах конвейера имеют одинаковый вид и отличаются лишь задержкой.

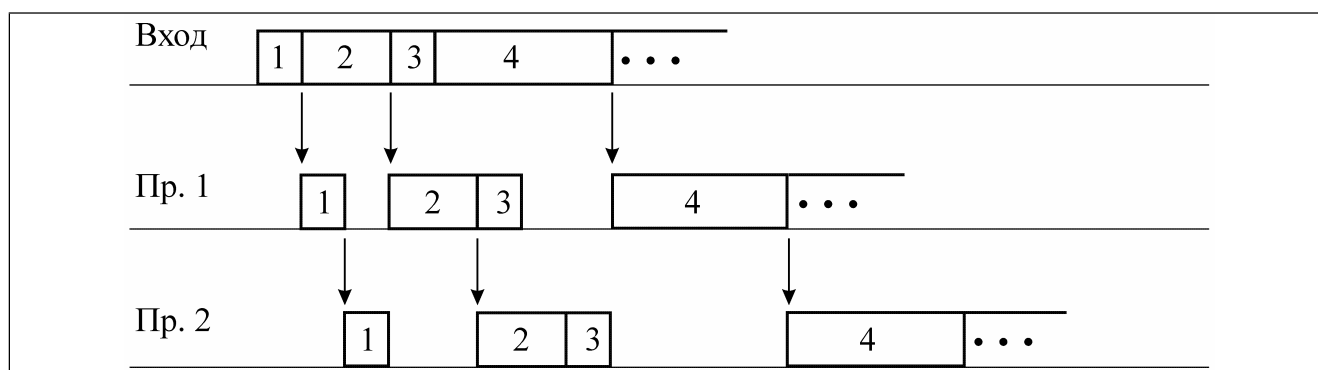


Рис. 5. Начало работы асинхронного конвейера

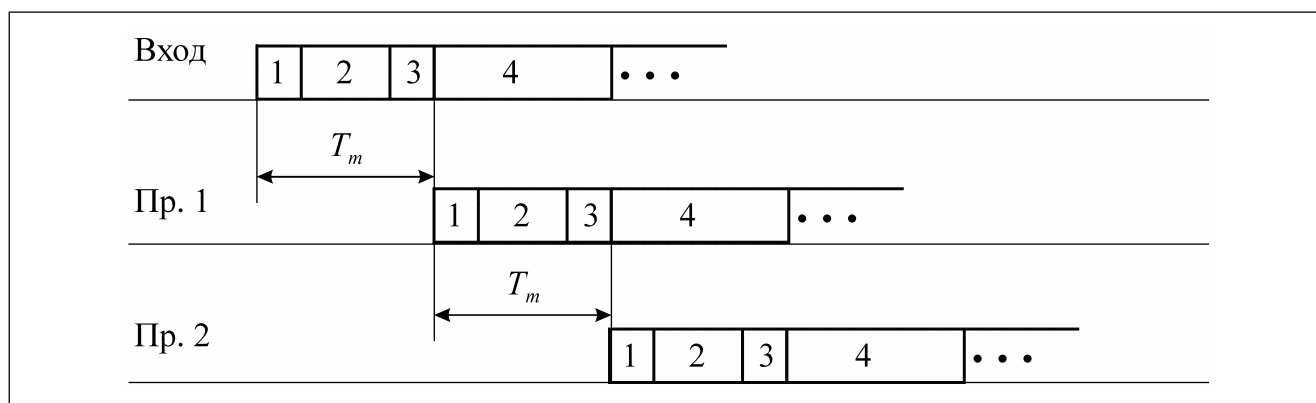


Рис. 6. Начало работы синхронного конвейера

Аналогичный эффект наблюдается и при наличии на входе конвейера свободных от работы тактов. В синхронном конвейере они передаются на выход без изменений с задержкой nT_m , где n – число этапов конвейера. В асинхронном конвейере свободные интервалы могут привести к сокращению задержки для последующих задач (рис. 7). При этом для со-

кращения задержки с nT_1 до nT_2 необходимо, чтобы свободные интервалы на входе имели суммарную длительность не менее $n(T_1 - T_2)$, $T_2 < T_1$, однако задачи между ними должны располагаться таким образом, чтобы позволить уменьшить задержку на каждом этапе на $T_1 - T_2$ (рис. 8).

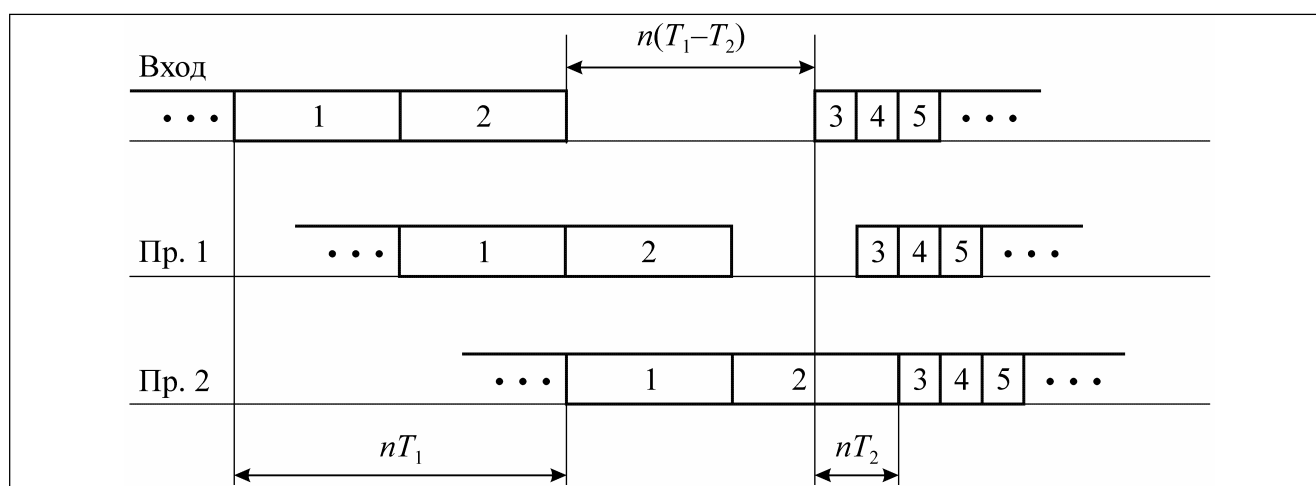


Рис. 7. Сокращение задержки в асинхронном конвейере за счет свободного интервала

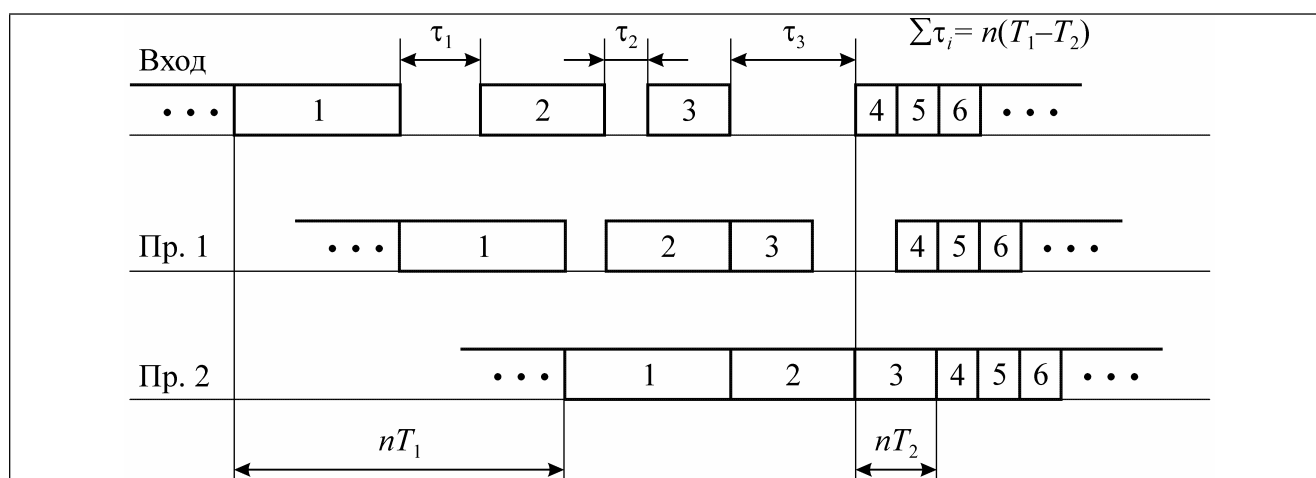
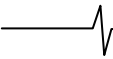


Рис. 8. Сокращение задержки в асинхронном конвейере за счет нескольких свободных интервалов

Таким образом, в асинхронном конвейере автоматически минимизируется время прохождения ин-

формации по конвейеру. Однако при этом для данной задачи это время зависит от всей предыстории



конвейера и поэтому трудно прогнозируемо. Наоборот, для синхронного конвейера время прохождения любой задачи фиксировано и равно nT_m , поэтому всегда известно, когда результаты любой задачи появятся на выходе конвейера.

В некоторых приложениях для сокращения задержки в синхронном конвейере бывает необходимо перейти от работы с максимальным тактом T_{m1} к работе с максимальным тактом T_{m2} , $T_{m2} < T_{m1}$. Временная диаграмма такого перехода аналогична временной диаграмме уменьшения задержки в асинхронном конвейере (рис. 7) и требует на входе свободного от работы интервала длительностью не менее $n(T_{m1} - T_{m2})$. Обратный переход от T_{m2} к T_{m1} , $T_{m2} < T_{m1}$, затруднений не вызывает, но приводит к прогрессивно увеличивающемуся вдоль конвейера простоя процессоров (аналогично временной диаграмме рис. 5).

Модификации базовой модели и алгоритмов работы БЗУ

До сих пор обсуждалась модель конвейера, описанная ранее условиями У1–У4. Рассмотрим, к чему приводят небольшие изменения этих условий.

а) Условие У3 о строгой пропорциональности объема информации длительности такта довольно жестко. В случае, когда оно не выполняется, введем для всех задач, обрабатываемых конвейером, коэффициенты

$$k_{ij} = v_{ij} / T_{ij}, i=1,2,\dots,n; j=1,2,\dots \quad (4)$$

где i – номер этапа, j – номер задачи, v_{ij} – объем входной информации i -того этапа для j -той задачи, T_{ij} – время обработки процессором i -того этапа этой задачи. Выберем для каждого этапа максимальный коэффициент

$$k_{im} = \max_j k_{ij}, i=1,2,\dots,n. \quad (5)$$

Тогда, как нетрудно проверить, все ранее изложенное остается справедливым и для рассматриваемого случая при замене k_i на k_{im} . При этом, правда, объем i -го БЗУ будет частично недоиспользоваться при обработке задачи с $k_{ij} < k_{im}$.

б) Рассмотренные структуры макроконвейера работоспособны и при отказе от точного равенства времен обработки фиксированной задачи на всех этапах конвейера (условие У2). Все представленные выводы справедливы и для случая, когда время обработки любых этапов конвейера может быть меньше длительности входного такта.

При этом для синхронного конвейера, по существу, ничего не меняется, за исключением некоторого простоя процессоров на тех этапах, где время обработки меньше, а для асинхронного конвейера может

наблюдаться уменьшение времени задержки последующих задач, аналогичное таковому при появлении свободных интервалов. Подобное ускорение прохождения задач по конвейеру может привести к увеличению необходимого объема БЗУ, так что при расчете коэффициентов k_{im} по формулам (4), (5) за величину T_{ij} необходимо принимать истинное время решения j -ой задачи, в то время как для синхронного конвейера величину T_{ij} достаточно приравнять величине входного такта для этой задачи. В результате объем БЗУ асинхронного конвейера может существенно возрасти, что является его недостатком. Устранение этого недостатка может быть произведено путем отказа от возможности обращения к БЗУ по записи в любой момент времени и введения разрешения на запись в БЗУ из процессора только при наличии достаточного объема свободной памяти для размещения результатов обработки очередной задачи. Однако такой метод приводит к усложнению способа управления конвейером.

в) Отказ от второй части условия У4 приводит к возможности считывания из БЗУ в процессор сразу нескольких задач. Временная диаграмма работы конвейера при этом не изменяется, если каждую такую «пачку» задач рассматривать как укрупненную макрозадачу. Такой режим работы может в некоторых случаях привести к уменьшению необходимого объема БЗУ. Например, если все возможные длительности тактов кратны некоторой величине δ , а за время δ из БЗУ успевает считываться информация любого такта, то при банковской организации БЗУ i -го этапа достаточно иметь 2 банка объемом $k_i T_m$. Однако такие случаи довольно специфичны и поэтому подробнее в данной работе рассматриваться не будут.

Закключение

Таким образом, два простых рассмотренных выше способа организации макроконвейера показывают, что работа с переменным тактом требует некоторого увеличения объема БЗУ в сравнении с макроконвейером с фиксированным тактом. Однако это увеличение не слишком велико (не более, чем в 1,5 раза) и является вполне приемлемой платой за минимизацию требований к производительности процессоров и получаемую гибкость устройства. В связи с этим подобные макроконвейеры должны найти широкое применение в различных областях использования вычислительной техники и, в частности, в трактах обработки сигналов многофункциональных РЛС.

УДК 621.372.542

ПРОГРАММА ДЛЯ РАСЧЕТА ЭКСТРЕМАЛЬНЫХ ПАРАМЕТРОВ ЦИФРОВЫХ И АНАЛОГОВЫХ ФИЛЬТРОВ И ЕЕ ПРИМЕНЕНИЕ

Д. В. Алёшин, А. Т. Мингазин

Введение

Минимально необходимый порядок N частотных цифровых или аналоговых фильтров, таких как фильтры нижних и верхних частот (ФНЧ и ФВЧ), полосовые и режекторные фильтры (ПФ и РФ) определяется заданными допустимыми исходными параметрами к их характеристикам ослабления. В [1] введены области допустимых исходных параметров S , вид и размерность которых зависит от типа фильтра (ФНЧ, ПФ или др.) и используемой аппроксимации его характеристики. Расчет фильтра для любой точки области параметров S будет приводить к допустимой характеристике. Выбор точки выполняется по тем или иным соображениям разработчиком. С увеличением N при фиксированных исходных параметрах область S расширяется, что дает возможность получения допустимых характеристик в широких пределах.

Итак, перед разработчиком стоит задача выбора исходных значений параметров в S для расчета собственно фильтра. В инженерной практике часто стремятся определить экстремально улучшенное значение одного из параметров при заданных остальных [2-6]. Например, неравномерность в полосе пропускания может быть экстремально уменьшена или полоса пропускания экстремально расширена без нарушения требований к характеристике. Такой подход означает, что совокупность исходных значений параметров будет принадлежать границе S .

Расчетные соотношения для определения ряда экстремальных параметров цифровых и аналоговых ФНЧ, ФВЧ, ПФ и РФ с характеристиками Баттерворта, Чебышева и Золотарева-Кауэра представлены в [3]. Еще два других варианта расчета для ФНЧ были предложены в [4], а позднее описаны в [5].

Вычисление экстремальных параметров вручную на калькуляторе утомителен, а в некоторых случаях не возможен из-за громоздкости соотношений, наличия численных процедур и неизбежности совершения ошибок. В данной статье, после необходимых определений, представлена удобная компьютерная программа Extremal фирмы «РАДИС Лтд» для расчета экстремальных параметров цифровых и аналоговых ФНЧ, ФВЧ, ПФ и РФ на основе соотношений из [3,4].

Представлена программа Extremal фирмы «РАДИС Лтд» для расчета экстремальных параметров цифровых и аналоговых фильтров нижних и верхних частот, полосовых и режекторных фильтров с характеристиками Баттерворта, Чебышева и Золотарева-Кауэра. Приведены примеры использования программы, иллюстрирующие интересные альтернативные значения параметров конкретных фильтров. Оригинальная программа Extremal является, полезным и удобным вспомогательным средством для получения желаемых характеристик различных фильтров и может найти широкое применение в инженерной и исследовательской работе.

Исходные параметры

Под исходными параметрами характеристики ослабления фильтров (ФНЧ, ФВЧ, ПФ и РФ) понимают:

- неравномерность в полосе пропускания Δa в дБ;
- минимальное ослабление в полосе задерживания a_0 в дБ;
- граничные частоты полос пропускания и задерживания f_1, f_2, f_3, f_4 , $f_i < f_{i+1}$.

Предполагается, что в полосе пропускания характеристика имеет максимальный уровень 0 дБ. Для ФНЧ и ФВЧ необходимо задать только две граничными частоты.

Задавшись значениями параметров $\Delta a = \Delta a_{\max}$, $a_0 = a_{0\min}$, $f_i = f_{in}$, $i=1,2,3,4$ и частотой дискретизации f_s , в случае цифровых фильтров, определяют N для выбранного типа фильтра. Для расчета собственно фильтра можно использовать эти значения исходных параметров, или любые другие, находящиеся внутри или на границе области S . Такой выбор параметров обеспечивает заданные требования к фильтру $\Delta a \leq \Delta a_{\max}$, $a_0 \leq a_{0\min}$ при номинальных граничных частотах f_{in} , $i=1,2,3,4$. Заметим, что знак равенства в этих неравенствах означает, что область S является точкой и улучшение параметров возможно только при увеличении N на 1, 2, ... в случае ФНЧ (ФВЧ) или на 2, 4, ... в случае ПФ (РФ).

Программа Extremal позволяет определить целый ряд альтернативных значений исходных параметров, которые могут быть использованы для дальнейшего расчета фильтра с улучшенными характеристиками. Перед представлением Extremal определим желаемые свойства характеристик с экстремальными параметрами.

Экстремальные параметры

Для ФНЧ и ФВЧ определим 7, а для ПФ и РФ 11 свойств характеристики ослабления с экстремальными параметрами, которые могут представлять

практический интерес.

Свойства вариантов характеристик ФНЧ и ФВЧ:

1. Минимум максимальной взвешенной ошибки;
2. Минимальная неравномерность в полосе пропускания;
3. Максимум минимального ослабления в полосе задерживания;
4. Максимально широкая полоса пропускания;
5. Максимально узкая переходная полоса при номинальной полосе пропускания;
6. Минимальная неравномерность не во всей полосе пропускания;
7. Максимум минимального ослабления не во всей полосе задерживания.

Свойство 1 означает, что взвешенные уровни пульсаций характеристики ослабления в полосе пропускания и задерживания равны. Это означает, что $(1 - 10^{-\Delta a / 20}) / (1 - 10^{-\Delta a_{\max} / 20}) = 10^{(a_{0\min} - a_0) / 20}$. (1)

Свойства 2...5 и 6...7 описаны в [3] и [4], соответственно.

Для фильтров инверсных Чебышева и Баттерворта свойство 6, а для фильтров Чебышева и Баттерворта свойство 7 не имеют смысла [1,4]. Возможность получения свойств 6 и 7 зависит от исходных требований к характеристике фильтра и значения N.

Свойства вариантов характеристик ПФ и РФ:

1. Минимум максимальной взвешенной ошибки;
2. Минимальная неравномерность в полосе пропускания;
3. Максимум минимального ослабления в полосе задерживания;
4. Максимально узкие переходные полосы при номинальной полосе пропускания;
5. Максимальное расширение полосы пропускания слева;
6. Максимальное расширение полосы пропускания справа;
7. Максимальное симметричное расширение полосы пропускания;
8. Минимальная неравномерность в правой части полосы пропускания;
9. Минимальная неравномерность в левой части полосы пропускания;
10. Минимальная неравномерность в средней части полосы пропускания ПФ и в полосах пропускания РФ за исключением областей примыкающих к переходным зонам;
11. Максимум минимального ослабления не во всей полосе задерживания.

Для фильтров Чебышева и Баттерворта свойство 11, для фильтров инверсных Чебышева и Баттерворта свойства 8...10, не имеют смысла. Возможность получения свойств 8...11 зависит от исходных требований к характеристике и значения N. В качестве примера на рис.1 представлены все 11 вариан-

тов характеристик ослабления $a(f)$ для ПФ. Там же показаны исходные и допустимые параметры. Все характеристики $a(f)$, за исключением варианта 1, соответствуют расчетной точке на границе области S. Точка для варианта 1 находится внутри S.

Программа Extremal

Программа Extremal выполняет расчет экстремальных параметров цифровых и аналоговых ФНЧ, ФВЧ, ПФ и РФ с характеристиками Баттерворта, Чебышева и Золотарева-Кауэра. Программа позволяет рассчитать экстремальные параметры для 7-ми вариантов характеристик ФНЧ и ФВЧ и 11-ти вариантов характеристик ПФ и РФ, упомянутых выше. В программе определяется минимальный порядок фильтра и допускается его изменение в сторону увеличения и обратно при одновременном наблюдении изменения параметров характеристик.

Кроме того, в Extremal имеется возможность работать в одной из четырех систем параметров для описания неравномерности в полосе пропускания и ослабления в полосе задерживания:

$$\{\Delta a, a_0\}, \{\delta_1, \delta_2\}, \{\delta_p, \delta_s\}, \{\varepsilon_1, \varepsilon_2\}.$$

Связь между системами параметров описывается как

$$\delta_1 = (1 - 10^{-\Delta a / 20}) / (1 + 10^{-\Delta a / 20}),$$

$$\delta_2 = 2(10^{-a_0 / 20}) / (1 + 10^{-\Delta a / 20});$$

$$\delta_p = 1 - 10^{-\Delta a / 20}, \quad \delta_s = 10^{-a_0 / 20};$$

$$\varepsilon_1^2 = 10^{\Delta a / 10} - 1, \quad \varepsilon_2^2 = 10^{a_0 / 10} - 1.$$

Переход от параметров какой-либо системы к параметрам другой и обратно выполняется по желанию пользователя до или после задания конкретных значений.

Основное окно программы с результатами расчета экстремальных параметров представлено на рис.2. Три выпадающих окна позволяют выбрать любой из упомянутых фильтров.

С помощью кнопки «Исходные параметры» вызывается вспомогательное окно необходимое для ввода исходных параметров и значения f_s . В данном случае $f_s = 1$. Исходные (см. исходный вариант) и рассчитанные экстремальные параметры (см. варианты 1...11), а также минимальный порядок фильтра N отображаются в поле основного окна. Для удобства исходные значения можно изменять непосредственно в строке «исходный».

Ниже представлены три примера расчета с помощью Extremal. В первом примере даны комментарии к результатам на рис.2. Еще в двух примерах иллюстрируются другие важные свойства фильтров с экстремальными параметрами.

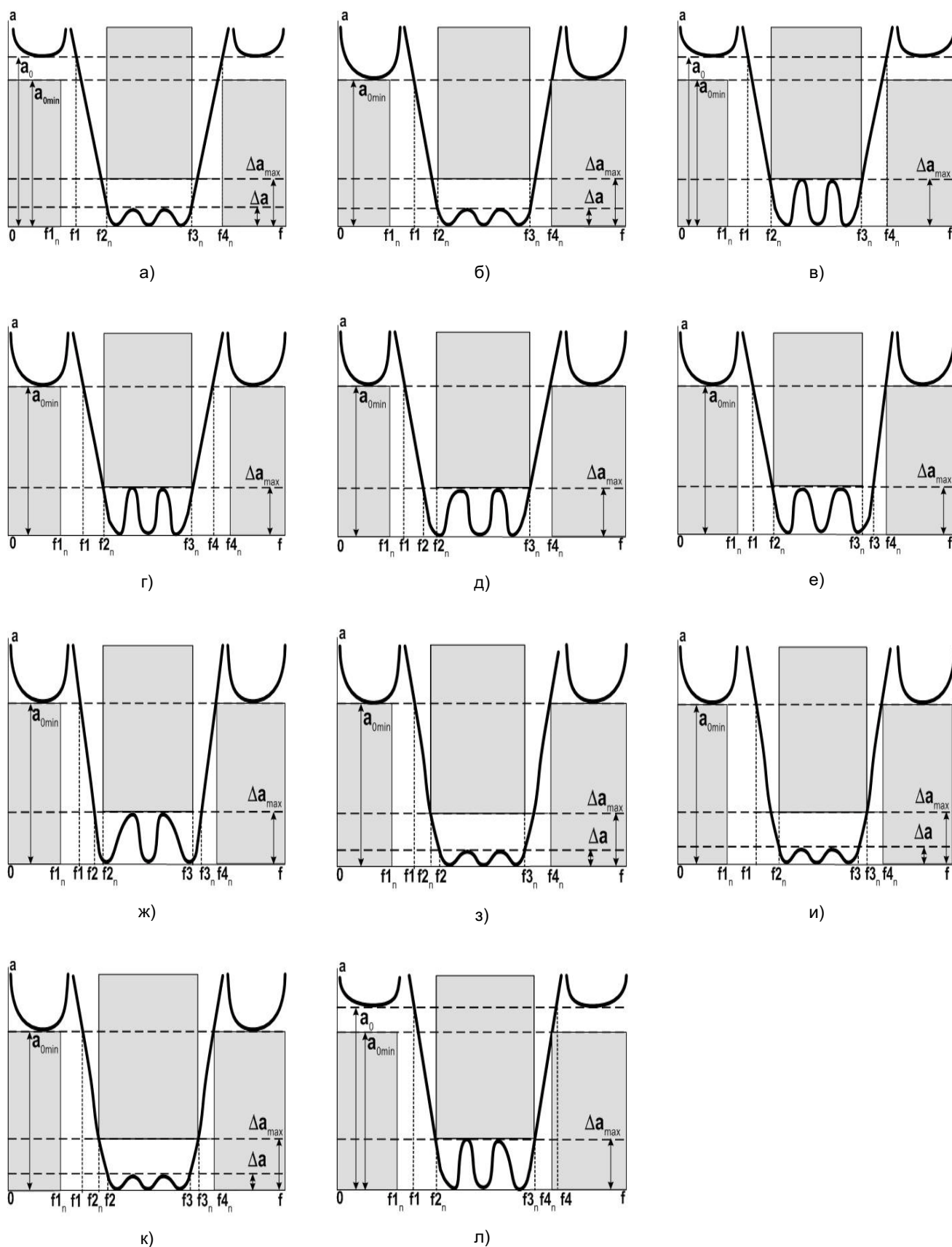


Рис.1. Характеристики ослабления ПФ со свойствами 1(а), 2(б), ..., 11(л).

Extremal						
Тип фильтра: Цифровой ПФ Золотарева-Кауэра Помощь						
Система параметров dA, A0 Иходные параметры N 8						
Вариант	dA, дБ	A0, дБ	F1	F2	F3	F4
Исходный	2	70	0.05	0.2	0.3	0.38
1	1.4523735	72.5142835	0.12	0.2	0.3	0.38
2	0.8728094	70	0.12	0.2	0.3	0.38
3	2	74.1957034	0.12	0.2	0.3	0.38
4	2	70	0.1290847	0.2	0.3	0.3709153
5	2	70	0.1052515	0.1789253	0.3	0.38
6	2	70	0.1256266	0.2	0.3076124	0.38
7	2	70	0.12	0.194407	0.305593	0.38
8	0.8040264	70	0.1212448	0.2017253	0.3	0.38
9	0.6847411	70	0.1185234	0.2	0.297934	0.38
10	0.5876428	70	0.12	0.2023467	0.2976533	0.38
11	2	76.0603401	0.11594	0.2	0.3	0.38406

Рис.2. Основное окно программы Extremal с результатами расчетов.

Примеры расчета

Пример 1. Как видим на рис.2. представлены рассчитанные экстремальные параметры и порядок цифрового ПФ Золотарева - Кауэра для следующих исходных данных:

$$\Delta a = \Delta a_{\max} = 2 \text{ дБ}; a_0 = a_{0\min} = 70 \text{ дБ};$$

$$f_1 = f_{1n} = 0,05; f_2 = f_{2n} = 0,2;$$

$$f_3 = f_{3n} = 0,3; f_4 = f_{4n} = 0,38; f_s = 1.$$

Частоты $f_1 \dots f_4$ на рис.2 обозначены как F1...F4, а параметры Δa и a_0 , как dA и A0.

Как видим на рис.2 порядок ПФ N=8. Исходный вариант соответствует исходным данным. Вариант 1 соответствует характеристике с минимумом максимальной ошибки и можно убедиться, что в этом случае справедливо соотношение (1). Вариант 2 дает характеристику с минимальной $\Delta a = 0,872809$ дБ в номинальной полосе пропускания, а вариант 3 дает максимум $a_0 = 74,196$ дБ в номинальной полосе задерживания (точнее в каждой из двух полос, т.к. это ПФ). Характеристике для варианта 4 соответствуют максимально узкие переходные полосы при номинальной полосе пропускания. Для варианта 5 полоса пропускания максимально расширена слева, для варианта 6 – справа, а для варианта 7 – симметрично. Варианту 8 соответствует характеристика с минимальной $\Delta a = 0,804026$ дБ, но лишь в правой части полосы пропускания от 0,201725 до 0,3 с монотонным увеличением неравномерности до 2 дБ в диапазоне частот от 0,201725 до 0,2. Аналогично можно трактовать варианты 9 и 10, но для левой и средней

части полосы пропускания с минимальными $\Delta a = 0,6847411$ дБ и $\Delta a = 0,5876428$ дБ. Наконец варианту 11 свойственна характеристика с максимумом $a_0 = 76,06$ дБ, но не во всей полосе задерживания. Это значение a_0 имеет место во всей заданной полосе задерживания слева и в полосе от 0,38406 до 0,5. Между частотами 0,38406 и 0,38 характеристика ослабления монотонно уменьшается от 76,06 дБ до 70 дБ. Качественные характеристики на рис.1а-л наглядно поясняют смысл всех полученных результатов для данного ПФ.

Следует заметить, что характеристики цифровых или аналоговых ПФ и РФ обладают определенной симметрией и если граничные частоты заданы без учета этого, то левый или правый склон характеристики ослабления будет иметь более узкую переходную полосу. Для обсуждаемого ПФ, а также для характеристик на рис.1 это левый склон. В программе Extremal этот факт учитывается и поэтому все значения частот $f_1 > f_{1n} = 0,05$ (см. рис. 2).

Для расчета собственно ПФ может быть выбран любой из 11 вариантов на рис.2. Очевидно, что при фиксированном N значение какого-либо параметра для вариантов 1...11 нельзя улучшить без ухудшения других параметров.

Пример 2. Рассмотрим расчет экстремальных параметров цифрового ФНЧ Золотарева - Кауэра со следующими исходными данными:

$$\Delta a = \Delta a_{\max} = 3 \text{ дБ}; a_0 = a_{0\min} = 45 \text{ дБ};$$

$$f_1 = f_{1n} = 0,1; f_2 = f_{2n} = 0,2; f_s = 1.$$

Минимальный порядок ФНЧ N=4. Вместе с экс-

тремальными параметрами, полученными с помощью Extremal, в табл.1 приведены соответствующие им другие данные о фильтрах. Здесь $\Delta\tau$ - неравномерность, а T_m - максимум характеристики ГВЗ в номинальной полосе пропускания; G - усиление шума округления. Значения ГВЗ даны в отсчетах частоты дискретизации.

Таблица 1

Вариант	$\Delta\alpha$, дБ	a_0 , дБ	f_1	f_2	$\Delta\tau$	T_m	G , дБ
1	0,810911	55,308	0,1	0,2	10,5	14,4	14,6
2	0,082258	45	0,1	0,2	4,3	7,7	10,7
3	3	62,164	0,1	0,2	19,8	23,1	17,7
4	3	45	0,144544	0,2	3,5	5,4	13,1
5	3	45	0,1	0,143387	24,3	27,3	17,2
6	7,7e-6	45	0,033909	0,2	3,5	7,1	9,2
7	3	102,487	0,1	0,364614	17,7	21,2	18,0

Расчет параметров $\Delta\tau$, T_m и G проведен с помощью программы синтеза каскадных БИХ- фильтров DIFID [7] при следующих установках: минимизация усиления шума округления - включена, форма звеньев 2-го порядка - прямая, норма для масштабирования - L_∞ , коэффициенты не квантуются, масштабные множители вводятся путем изменения коэффициентов числителя передаточной функции, округление в звеньях выполняется на последнем этапе суммирования.

Результаты в табл.1 имеют значительный разброс. Как видим, минимум $\Delta\tau=3,5$ получен для 4 и 6-го вариантов, минимум $T_m=5,4$ - для 4-го варианта и минимум $G=9,2$ - для 6-го варианта.

Пример 3. Интересно, как изменятся параметры фильтра при увеличении его порядка без изменения других исходных данных. Для $N=5$ в табл. 2 представлены результаты расчетов.

Таблица 2

Вариант	$\Delta\alpha$, дБ	a_0 , дБ	f_1	f_2	$\Delta\tau$	T_m	G , дБ
1	0,195309	67,369	0,1	0,2	10,3	15,5	13,5
2	0,001158	45	0,1	0,2	2,2	5,3	8,1
3	3	80,721	0,1	0,2	31,7	36,3	18,7
4	3	45	0,172609	0,2	1,9	4,0	13,5
5	3	45	0,1	0,118863	53,7	57,5	18,8

Для данного случая совокупность исходных параметров при $N=5$ не позволяют получить характеристики для вариантов 6 и 7. Сравнивая данные в табл. 1 и табл.2, можно наблюдать улучшение параметров характеристик при переходе от $N=4$ к $N=5$.

Например, неравномерность $\Delta\alpha$ в номинальной полосе пропускания можно уменьшить от 0,082258 дБ до 0,001158 дБ, а неравномерность $\Delta\tau$ - от 3,5 до 1,9. В тоже время при таком переходе для варианта 5 имеет место ухудшение параметров характеристики ГВЗ.

По приведенным результатам можно сделать и другие интересные наблюдения. Таким образом, программа Extremal предлагает разработчикам полезные альтернативы по расчету фильтров. Выбор того или иного варианта расчета зависит от требований к конкретному проекту.

Заключение

Программа Extremal позволяет гибко оперировать значениями исходных параметров и порядков различных цифровых и аналоговых фильтров с целью достижения желаемых характеристик и является полезным и удобным вспомогательным средством для разработчиков радиоэлектронной аппаратуры. Представляется, что Extremal целесообразно интегрировать в существующие или вновь создаваемые программы по проектированию цифровых и аналоговых фильтров.

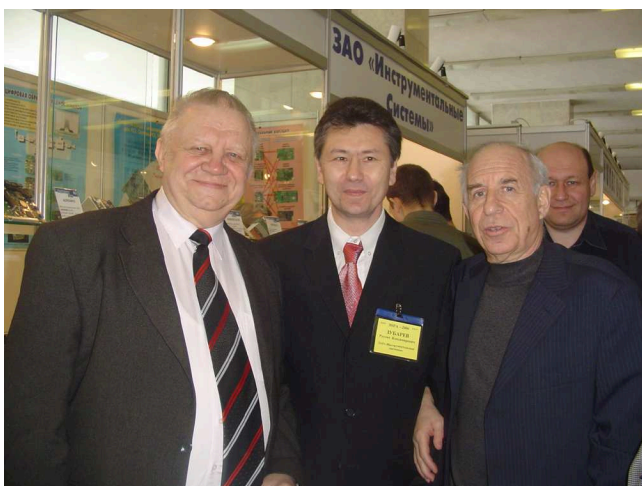
Литература

1. Мингазин А.Т. Начальные приближения для синтеза цифровых фильтров с минимальной длиной слова коэффициентов. // Электронная техника. 1983. Сер. 10. № 6. С. 3-8.
2. Савченко С.М., Смирнов Э.Е. Улучшение частотных характеристик эллиптических фильтров. // Изв. вузов СССР. Сер. Радиоэлектроника. 1976. Т. XIX. № 6. С.113-116.
3. Мингазин А.Т. Экстремальные параметры аналоговых и цифровых фильтров. // Электросвязь. 1999. № 1. С. 22-23.
4. Мингазин А.Т. Способ улучшения характеристик затухания фильтров. // Радиотехника. 1985. № 9. С. 89-91.
5. Tomic D.V., Lutovac M. D., Evans B.L. Advanced digital IIR filter design. // ECCTD. 1999. P.1323-1326.
6. Vlcek M., Unbehauen R. Degree, ripple and transition width of elliptic filters. // IEEE Trans. 1989. CAS-36. № 3. P.469-472.
7. Мингазин А.Т. Программа DIFID: эффективный синтез каскадных цифровых БИХ-фильтров. // IV Международная конференция «Цифровая обработка сигналов и ее применение»(DSPA). 2002. Т.1. Февраль-Март. С. 90-93.

ИНФОРМАЦИЯ

8-я Международная научно-техническая конференция и Выставка «Цифровая обработка сигналов и её применение – DSPA'2006»

Развитие современных информационных технологий характеризуется широким практическим использованием методов и техники цифровой обработки сигналов (ЦОС) - одной из самых динамичных и быстро развивающихся технологий в мире телекоммуникаций и информатизации общества.



Отвечая целям и задачам консолидации и поддержки отечественных ученых и специалистов, работающих в области передовых информационных технологий реального времени, расширения международных научных связей и содействия внедрению высоких наукоёмких технологий XXI века, Российское НТОРЭС имени А.С. Попова совместно с компанией «AUTEX.Ltd» при поддержке РФФИ и при участии Федерального агентства по промышленности РФ, Министерства образования и науки РФ, Российской секции IEEE Signal Processing Society, Института радиотехники и электроники РАН, Института проблем управления РАН, Института проблем передачи информации РАН, МГТУ им. Н.Э. Баумана, МТУСИ, МАИ, МЭИ и других ведущих российских вузов и организаций провели 29 – 31 марта 2006 г. в ИПУ РАН, г. Москва, 8-ю Международную научно-техническую конференцию и выставку «Цифровая обработка сигналов и её применение - DSPA'2006». В работе конференции и выставки приняли участие около 1500 человек из 34 городов России и стран СНГ, а также ученые из США, Германии, Финляндии, Польши, Тайланда, Пакистана, Ирана.

Работа конференции проходила по десяти секциям: «Теория сигналов и систем» (руководители – д.т.н., проф. Ю.С. Шинаков, д.т.н., проф. М.С. Немировский); «Теория и методы цифровой обработки сигналов» (руководитель – д.т.н., проф. Ю.А. Брюханов); «Обработка сигналов в системах телекоммуникаций» (руководитель – д.т.н., проф. Ю.Н. Прохоров); «Обработка и передача изображений» (руководители – д.т.н. В.П. Дворкович, к.ф.-м.н. А.В. Дворкович); «Обработка сигналов в радиотехнических системах»

(руководители – д.т.н., проф. Сосулин Ю.Г., к.т.н., проф. Сперанский В.С.); «Обработка измерительной информации» (руководитель – к.т.н., проф. Скачко Ю.В.); «Цифровое радиовещание» (руководители – к.т.н., доц. Е.П. Зелевич, к.т.н., с.н.с. Щербина В.И.); «Цифровая обработка многомерных сигналов» (руководители – д.т.н., проф. Миронов В.Г., к.т.н., доц. Чобану М.К.); «Проектирование и техническая реализация систем ЦОС» (руководитель – д.т.н., проф. В.В. Витязев); «Нейрокомпьютерная обработка сигналов и изображений» (руководитель – д.т.н., проф. Галушкин А.И.). На конференции было заслушано около 200 докладов, по материалам которых изданы «Труды Конференции» в 2-х томах, общим объемом – 667 стр.

В своем приветствии к участникам конференции Председатель Национального комитета член-корр. РАН Ю.Б. Зубарев отметил масштабность и значимость для России открывающегося научного форума, растущий из года в год интерес широкой аудитории его участников к проблемам цифровых информационных технологий реального времени, всё более нарастающий приток молодежи.

На пленарном заседании в день открытия конференции были заслушаны 6 обзорных докладов, в которых нашли отражение вопросы теории и практики построения систем цифровой обработки изображений и звука в самых различных приложениях. Основной акцент был сделан в направлении решения проблем мобильного узкополосного вещания и IP TV (Дворкович В.П., Дворкович А.В., Зубарев Ю.Б. и др. ФГУП «НИИ радио», г. Москва). С большим интересом встречен доклад представителей г. Рязани (Алпатов Б.А., Бабаян П.В., Костяшкин Л.Н.), посвященный технологиям обработки изображений в системах обнаружения и сопровождения объектов. О новых эффективных способах передискретизации изображений рассказал в своем выступлении проф. Сарамьяки Т. (Институт обработки сигналов Тамперского технологического университета, Финляндия). Методы и технологии обработки акустических сигналов для пространственного воспроизведения звука на основе волнового поля стали предметом обсуждений в докладе специалистов из Германии (Рабинштейн Р., Спорс С., Нюрнбергский университет, лаборатория телекоммуникаций). С новыми техническими решениями построения оптимальных декодеров на основе многопороговых алгоритмов для высокоскоростных спутниковых систем познакомил участников проф. Золотарев В.В. (содокладчик Зубарев Ю.Б., МНИТИ).

Одновременно с работой конференции компанией «AUTEX.Ltd» была организована выставка новых информационных технологий реального времени и инструментальных средств ЦОС, в работе которой приняли участие ведущие в этой области отечественные

предприятия и фирмы: ЗАО «AUTEX.Ltd», ЗАО «Инструментальные системы», компании «AVD Systems», «Alternative Solutions», ЗАО «L-Card», ООО «Растр технологий», НТЦ «Модуль», ГУП НПЦ «ЭЛВИС», ООО «Руднев-Шилиев», ООО «Центр АЦП», ООО «НПП Цифровые решения» и др.



В дни работы Конференции и Выставки состоялись семинары-презентации по новым разработкам в области DSP-технологий, которые провели хорошо известные на российском рынке фирмы: ГУП НПЦ «ЭЛВИС», ЗАО «Инструментальные системы», ЗАО «СПИРИТ» и др. Как и в предыдущие годы, наиболее представительным и посещаемым был специализированный семинар компаний «Analog Devices Inc.-AUTEX. Ltd» по DSP-продукции «Analog Devices», привлечший внимание около 500 участников.

С целью создания условий для всестороннего открытого обсуждения проблем и перспектив развития DSP-технологий и их широкого использования в разработках отечественных ученых и специалистов в первый день работы Конференции и Выставки был проведен «Круглый стол – DSPA» с участием представителей ведущих российских фирм и организаций.

В последние годы в рамках Конференции и Выставки все большее внимание уделяется поддержке молодых ученых, аспирантов и студентов. Проводятся конкурсы на лучший научный доклад, организуются семинары-презентации по университетским программам (в этом году прошел семинар-презентация «Университетская программа Texas Instruments»), активизируется работа соответствующей секции.

На заключительном заседании, которое открыл Председатель Национального оргкомитета, чл.-корр. РАН Ю.Б. Зубарев, после обсуждения итогов работы конференции и выработки соответствующего реше-

ния, состоялось награждение молодых ученых и специалистов за лучшие доклады по каждой из секций.

Следующая, 9-я Международная конференция и выставка «DSPA'2007», состоится ориентировочно в марте 2007 г. в г. Москва, ИПУ РАН. (следите за информацией на нашем сайте: www.dspra.ru).

Информационную поддержку Конференции и Выставки обеспечивали журналы: «В мире науки», «Датчики и системы», «Компоненты и технологии», «Радиотехника», «Современная электроника», «Схемотехника», «Цифровая обработка сигналов», «Chip News», «Электроника: НТБ», «Электронные компоненты», «Электросвязь»; издательство «Горячая линия – Телеком».

По материалам прошедшей Конференции редакции научно-технических журналов «Цифровая обработка сигналов», «Электросвязь» и «Радиотехника» планируют в 2006 году ряд тематических выпусков, и в первую очередь: «Цифровая обработка изображений» «Цифровая обработка сигналов в системах телекоммуникаций», «Цифровая обработка сигналов в радиотехнических системах».

