

УДК 621.391.266

ЭКОНОМИЧНЫЙ МЕТОД ОЧИСТКИ РЕЧИ ОТ ШУМА, ОСНОВАННЫЙ НА БЛОЧНОМ ПРЕДСТАВЛЕНИИ СИГНАЛА В ПРОСТРАНСТВЕ СОСТОЯНИЙ И ВЕКТОРНОМ КВАНТОВАНИИ

А. Я. КАЛЮЖНЫЙ*, В. Ю. СЕМЕНОВ**

*Научно-производственное предприятие “Дельта”, Киев

**Институт гидромеханики НАН Украины, Киев

Получено 01.08.2002

Предложен новый экономичный метод очистки речевых сигналов от шума. Сокращение вычислений основано на свойствах блочной модели авторегрессионного сигнала, развитой в работе. Показано уменьшение ошибки по сравнению с традиционным фильтром Калмана. Рассмотрена задача оценивания авторегрессионных параметров речи в присутствии шума. Предложена экономичная двухэтапная процедура оценивания, основанная на векторном квантовании авторегрессионных параметров. На первом этапе начальное приближение для оптимального набора авторегрессионных параметров определяется на небольшом количестве квантов. Затем значение оценки улучшается с помощью эффективной итерационной процедуры. Эффективность результирующего метода проверена на реальных речевых сигналах.

Запропоновано новий економічний метод очищення мовних сигналів від шуму. Скорочення обчислень базується на властивостях розробленої в роботі блокової моделі авторегресійного сигналу. Показано зменшення похибки у порівнянні з традиційним фільтром Калмана. Розглянуто задачу оцінювання авторегресійних параметрів мовних сигналів у присутності шуму. Запропоновано економічну двоетапну процедуру оцінювання, яка базується на векторному квантуванні авторегресійних параметрів. На першому етапі початкове наближення для оптимального набору авторегресійних параметрів визначається на невеликій кількості квантів. Потім значення оцінки поліпшується за допомогою ефективною ітераційною процедурою. Ефективність результируючого методу перевірена на реальних мовних сигналах.

A computationally efficient method for enhancement of the speech signals is proposed. Reduction of computations is obtained due to properties of the developed block model of the autoregressive signal. A decrease of the filtering error in comparison to the traditional Kalman filter is shown. The problem of estimation of the speech autoregressive parameters at noise background is considered. A two-phase computationally efficient estimation procedure, based on the vector quantization of the autoregressive parameters, is proposed. At the first stage the initial approximation for the optimal set of autoregressive parameters is determined on small number of the autoregressive quantums. Then the value of estimate is improved by efficient iterative procedure. The performance of the resulting method is evaluated on real speech signals.

ВВЕДЕНИЕ

Задача очистки речевых сигналов от шума имеет большое практическое значение. Общее качество речи и ее разборчивость могут существенно ухудшаться в присутствии фонового шума, особенно когда сигнал подвергается дальнейшей обработке [1–4]. Речевые кодеки, устройства распознавания речи и идентификации диктора, построенные для обработки незашумленных сигналов, оказываются при наличии фоновых шумов малоэффективными или вообще бесполезными.

Большинство современных методов обработки речевых сигналов основаны на использовании авторегрессионной модели речеобразования [1–6], в которой речевой сигнал представляется, как результат прохождения управляющего (возбуждающего) процесса через авторегрессионный (полюсной) фильтр. Суть данной модели отображена на рис. 1.

Возбуждающий процесс моделирует поток воз-

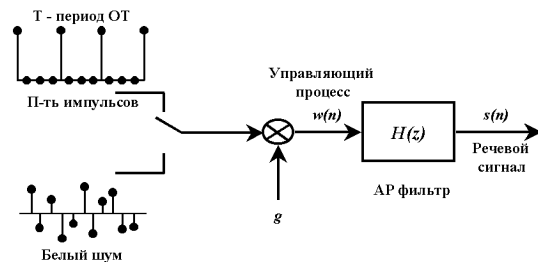


Рис. 1. Блок-схема авторегрессионной модели речеобразования

духа на выходе голосовых связок человека. В данной модели он представляет собой белый шум в случае произнесения глухих звуков (невокализованная речь) или последовательность импульсов, идущих друг за другом с периодом, соответствующим частоте колебания голосовых связок (вокализованная речь). В случае вокализованной речи период следования импульсов возбуждающего про-

песа называется периодом основного тона. Его значение может меняться от 2 мс (у женщин и детей) до 20 мс (у мужчин с низким голосом).

Работа полюсного фильтра $H(z)$, моделирующего форму голосового тракта в момент произнесения звука, может быть описана в разностной форме:

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

где $s(n)$ – речевой сигнал; $w(n)$ – возбуждающий процесс; g – коэффициент усиления; a_k , $k=1, 2, \dots, p$ – авторегрессионные коэффициенты. Порядок авторегрессионной модели p выбирается, как правило, равным $8 \div 20$. Очевидно, что повышение порядка модели позволяет более точно моделировать спектральные характеристики речи.

Основным свойством параметров авторегрессионной модели является то, что они относительно медленно меняются с течением времени. Можно считать, что эти параметры остаются неизменными на отрезках длиной $10 \div 30$ мс (свойство квазистационарности речевых сигналов).

Традиционно авторегрессионные коэффициенты определяются с помощью методов линейного предсказания, в частности автокорреляционного метода [3–6]. Поэтому авторегрессионные коэффициенты также зачастую называют коэффициентами линейного предсказания. На рис. 2 показан пример аппроксимации спектра мощности речевого сигнала (продолжительностью 20 мс) с помощью функции

$$|H(e^{i\omega})|^2 = \frac{g^2}{\left|1 + \sum_{k=1}^p a_k e^{-ik\omega}\right|^2},$$

вычисленной посредством автокорреляционного метода.

Итак, предполагается, что доступны только результаты наблюдений

$$z(n) = s(n) + v(n), \quad (2)$$

где $v(n)$ – белый шум с дисперсией σ_v^2 . Наиболее распространенным способом фильтрации авторегрессионных последовательностей является фильтр Калмана [1, 2, 4, 7]. Его применение предполагает представление уравнений (1) и (2) в пространстве состояний:

$$\begin{cases} \mathbf{x}(n) = \mathbf{F}(n)\mathbf{x}(n-1) + \mathbf{G}(n)w(n), \\ z(n) = \mathbf{C}\mathbf{x}(n) + v(n). \end{cases} \quad (3)$$

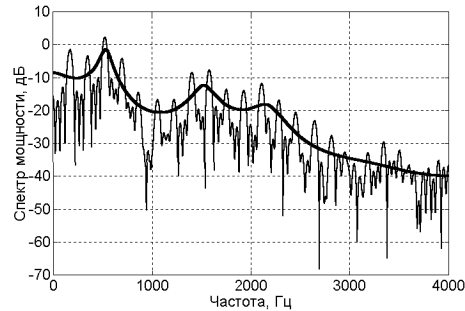


Рис. 2. Пример аппроксимации спектра мощности речевого сигнала

Матрицы \mathbf{F} и \mathbf{G} определяются через параметры уравнения (1), а вектором состояния является [1, 2, 4]

$$\mathbf{x}(n) = [s(n-p+1), s(n-p+2), \dots, s(n-1), s(n)]^T.$$

Отметим, что в модели (3) два последовательных вектора состояния $\mathbf{x}(n)$ и $\mathbf{x}(n+1)$ отличаются друг от друга только на два дискретных отсчета.

Применение фильтра Калмана к системе (3) для вычисления оценки $\hat{\mathbf{x}}(n) = E\{\mathbf{x}(n)|z(n), \dots, z(1)\}$ связано со значительными вычислительными затратами, в первую очередь, потому что на каждом временном шаге, совпадающем с тактом дискретизации, обновляются все внутренние переменные. Несмотря на разреженную структуру матриц модели (3), это требует большого количества операций (формулы для его оценки приводятся, например, в [1]). Кроме того, необходимо отметить, что обновление внутренних переменных калмановского алгоритма при поступлении каждого очередного наблюдения находится в некотором противоречии с процедурой оценивания параметров, которая обычно выполняется на блоках данных.

С учетом сказанного, в данной работе предлагается новый подход к фильтрации зашумленных авторегрессионных сигналов, основанный на их блочном представлении. Построенная блочная модель подразумевает обработку сигнала неперекрывающимися блоками произвольной длины. Будет показано, что фильтр Калмана, основанный на таком представлении, обеспечивает существенное сокращение вычислительных затрат, по сравнению с традиционным фильтром Калмана, и значительное снижение ошибки фильтрации.

Другая принципиальная проблема существующих методов очистки речи от шума связана с отсутствием надежных алгоритмов оценивания параметров авторегрессионного сигнала в присутствии шумов. Наиболее популярный подход к ре-

шению этой задачи состоит в максимизации функционалов правдоподобия или апостериорной вероятности с помощью алгоритма EM (estimate-maximize) [2, 3]. Однако известные методы (имеющие, как правило, итерационную структуру) не обеспечивают глобальной максимизации и сильно зависят от выбора начального приближения. Поэтому в работе [4] был предложен помехоустойчивый метод оценивания авторегрессионных параметров, основанный на векторном квантовании. Значение оценки определялось путем максимизации функционала правдоподобия (зависящего от параметров помехи) на ограниченном количестве авторегрессионных квантов. Основное препятствие для применения этой методологии состоит в высоких вычислительных затратах, поскольку необходимый размер кодовой книги (количество квантов) может быть довольно большим. В данной статье показано, как уменьшить вычислительную нагрузку этого подхода с помощью свойств блочной модели авторегрессионного сигнала.

Другой путь к снижению вычислительных затрат состоит в применении следующего двухэтапного алгоритма. На первом этапе для поиска начального приближения используется прямая максимизация функционала правдоподобия на небольшом количестве квантов. Затем точное значение оценки определяется с помощью хорошо известной итерационной процедуры [2], которая обеспечивает глобальную максимизацию функционала правдоподобия при условии качественной инициализации алгоритма. Рассмотрена вычислительная оптимизация указанной итерационной процедуры. Полученный комбинированный метод эффективно решает важную проблему итерационных методов, связанную с выбором начального приближения.

1. БЛОЧНОЕ ПРЕДСТАВЛЕНИЕ АВТОРЕГРЕССИОННОГО СИГНАЛА

В этом разделе описывается блочное представление моделей (1) и (2), а также приводятся его свойства, обеспечивающие сокращение вычислительных затрат.

Введем следующий сигнальный вектор:

$$\mathbf{S}(k) = [s((k-1)l+1), s((k-1)l+2), \dots, s(kl)]^T. \quad (4)$$

В отличие от стандартного представления в пространстве состояний (3), вектор (4) содержит произвольное число l элементов ($l \geq p$). Кроме того, смежные по времени вектора не перекрываются. Аналогично представлению (4), введем на k -ом временном блоке векторы возбуждения, шума

и наблюдений: $\mathbf{W}(k)$, $\mathbf{V}(k)$, $\mathbf{Z}(k)$. Тогда формулы (1) и (2) могут быть представлены в следующем блочном виде (предполагается, что длина блока не превосходит период квазистационарности речи):

$$\begin{cases} \mathbf{AS}(k) = g\mathbf{W}(k), \\ \mathbf{Z}(k) = \mathbf{S}(k) + \mathbf{V}(k), \end{cases} \quad (5)$$

где \mathbf{A} – отбеливающая матрица с элементами

$$\mathbf{A}_{ij} = \begin{cases} 1, & i = j, \\ a_{i-j}, & j < i \leq j + p, \\ 0, & i < j \mid i > j + p. \end{cases} \quad (6)$$

Из системы (5) получаем вектор наблюдений:

$$\mathbf{Z}(k) = g\mathbf{A}^{-1}\mathbf{W}(k) + \mathbf{V}(k). \quad (7)$$

Если учитывать зависимость от предыдущего временного блока, выражение (7) может быть заменено системой в пространстве состояний (здесь и далее для простоты обозначений опускаются индексы, указывающие на зависимость матриц \mathbf{A} и \mathbf{F} от времени):

$$\begin{cases} \mathbf{S}(k) = \mathbf{FS}(k-1) + g\mathbf{A}^{-1}\mathbf{W}(k), \\ \mathbf{Z}(k) = \mathbf{S}(k) + \mathbf{V}(k). \end{cases} \quad (8)$$

В матрице \mathbf{F} первые $l-p$ столбцов являются нулевыми. Рекуррентные соотношения, связывающие остальные элементы матрицы \mathbf{F} с параметрами авторегрессионной модели (1), имеют вид

$$\mathbf{F}_{ij} = - \sum_{k=1}^{\min(i-1,p)} a_k \mathbf{F}_{i-k,j} + u_{ij}, \\ 1 \leq i \leq l, \quad l-p+1 \leq j \leq l,$$

где

$$u_{ij} = \begin{cases} a_{l+i-j}, & 1 \leq i \leq p-l+j, \\ 0 & \text{в других случаях.} \end{cases}$$

Основное свойство, связывающее матрицы \mathbf{A} и \mathbf{F} , состоит в особой структуре матрицы $\mathbf{T} = \mathbf{AF}$:

$$\mathbf{T}_{ij} = \begin{cases} -a_{l+i-j}, & 1 \leq i \leq p, \quad l-p+i \leq j \leq l, \\ 0 & \text{в других случаях.} \end{cases}$$

Таким образом, только $1+2+\dots+p = p(p+1)/2$ элементов матрицы \mathbf{AF} являются ненулевыми.

Например, при порядке $p=10$, длительности блока 25 мс и частоте дискретизации 8000 Гц отличны от нуля лишь 55 элементов из 40000. Более того, матрица \mathbf{AF} вообще не требует никаких вычислений, поскольку состоит из авторегрессионных коэффициентов, взятых с обратным знаком. В дальнейшем будем говорить, что некоторая матрица \mathbf{T} имеет АФ-структуру, если $T_{ij}=0, i > j-l+p$. Можно показать, что такие матрицы обладают следующими полезными свойствами.

1. Пусть \mathbf{C} – некоторая верхняя треугольная матрица. Тогда матрицы \mathbf{CT} и \mathbf{TC} имеют АФ-структуру.
2. Пусть \mathbf{C} – верхняя треугольная матрица. Тогда матрица $\mathbf{D}=\mathbf{TC}^{-1}$ имеет АФ-структуру.

Приведенные факты играют решающую роль для сокращения вычислений при фильтрации авторегрессионных процессов.

2. АЛГОРИТМ БЛОЧНОЙ ФИЛЬТРАЦИИ

Применим к системе (8) фильтр Калмана [7]:

$$\begin{aligned} \mathbf{P}(k|k-1) &= \mathbf{FP}(k-1|k-1)\mathbf{F}^T + \\ &\quad + g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T}, \\ \mathbf{P}(k|k) &= (\mathbf{P}^{-1}(k|k-1) + \sigma_v^{-2} \mathbf{E}_l)^{-1}, \\ \hat{\mathbf{S}}(k) &= \mathbf{F}\hat{\mathbf{S}}(k-1) + \sigma_v^{-2} \mathbf{P}(k|k)(\mathbf{Z}(k) - \\ &\quad - \mathbf{F}\hat{\mathbf{S}}(k-1)), \end{aligned} \quad (9)$$

где $\hat{\mathbf{S}}(k)$ – вектор оценок сигнала на k -ом временном блоке, полученный по блокам наблюдений $\mathbf{Z}(1), \dots, \mathbf{Z}(k)$; $\mathbf{P}(k|k)$ и $\mathbf{P}(k|k-1)$ – ковариационные матрицы ошибок фильтрации и предсказания соответственно; \mathbf{E}_l – единичная матрица размером $l \times l$. Вычисление ковариационных матриц $\mathbf{P}(k|k)$, $\mathbf{P}(k|k-1)$ и обновление оценок сигнала теперь выполняется только при поступлении нового блока измерений.

Выведем ключевое свойство блочного фильтра Калмана, состоящее в том, что обратная матрица ошибок $\mathbf{P}^{-1}(k|k)$ имеет $(2p+1)$ -диагональную структуру (т.е. все ее ненулевые элементы сосредоточены на $(2p+1)$ центральных диагоналях). Из выражения (9) можно показать, что

$$\begin{aligned} \mathbf{P}^{-1}(k|k) &= \mathbf{P}^{-1}(k|k-1) + \sigma_v^{-2} \mathbf{E}_l = \\ &= (\mathbf{FP}(k-1|k-1)\mathbf{F}^T + g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T})^{-1} + \\ &\quad + \sigma_v^{-2} \mathbf{E}_l. \end{aligned} \quad (10)$$

Перепишем (10), используя лемму об обращении матриц [8]:

$$\begin{aligned} \mathbf{P}^{-1}(k|k) &= -g^{-4} \mathbf{A}^T \mathbf{AF} \times \\ &\times [\mathbf{P}^{-1}(k-1|k-1) + g^{-2} \mathbf{F}^T \mathbf{A}^T \mathbf{AF}]^{-1} \times \\ &\times \mathbf{F}^T \mathbf{A}^T \mathbf{A} + \sigma_v^{-2} \mathbf{E}_l + g^{-2} \mathbf{A}^T \mathbf{A}. \end{aligned} \quad (11)$$

Учитывая разложение Холецкого [8]

$$\mathbf{P}^{-1}(k-1|k-1) + g^{-2} \mathbf{F}^T \mathbf{A}^T \mathbf{AF} = \mathbf{C}^T \mathbf{C}, \quad (12)$$

(\mathbf{C} – верхняя треугольная матрица), формула (11) примет вид

$$\begin{aligned} \mathbf{P}^{-1}(k|k) &= -g^{-4} \mathbf{A}^T \mathbf{AF} [\mathbf{C}^T \mathbf{C}]^{-1} \times \\ &\times \mathbf{F}^T \mathbf{A}^T \mathbf{A} + g^{-2} \mathbf{A}^T \mathbf{A} + \sigma_v^{-2} \mathbf{E}_l = \\ &= -g^{-4} \mathbf{A}^T \mathbf{AFC}^{-1} (\mathbf{A}^T \mathbf{AFC}^{-1})^T + \\ &\quad + g^{-2} \mathbf{A}^T \mathbf{A} + \sigma_v^{-2} \mathbf{E}_l. \end{aligned} \quad (13)$$

Рассмотрим слагаемое $\mathbf{T}=\mathbf{A}^T \mathbf{AFC}^{-1}$ в выражении (13). Матрица $\mathbf{D}=\mathbf{A}^T (\mathbf{AF})$ в соответствии со свойством 1 имеет АФ-структуру, так как \mathbf{A}^T является верхней треугольной матрицей. Тогда, согласно свойству 2, $\mathbf{T}=\mathbf{DC}^{-1}=\mathbf{A}^T \mathbf{AFC}^{-1}$ – также матрица с АФ-структурой (\mathbf{C}^{-1} – верхняя треугольная). Наконец, легко видеть, что матрица $\mathbf{TT}^T=\mathbf{A}^T \mathbf{AFC}^{-1} (\mathbf{A}^T \mathbf{AFC}^{-1})^T$ является $(2p+1)$ -диагональной.

Как видно из выражения (6), матрица \mathbf{A} имеет $p+1$ ненулевых центральных диагоналей; поэтому $\mathbf{A}^T \mathbf{A}$ является $(2p+1)$ -диагональной. Таким образом, все слагаемые в правой части (13) являются $(2p+1)$ -диагональными матрицами. Это завершает доказательство того, что $\mathbf{P}^{-1}(k|k) - (2p+1)$ -диагональная ($k \geq 1$), независимо от вида начальной матрицы $\mathbf{P}^{-1}(0|0)$. Из соотношений (9) следует, что такой же особенностью обладает и матрица $\mathbf{P}^{-1}(k|k-1)$. Отметим, что указанным свойством не обладают ковариационные матрицы стандартного фильтра Калмана для системы (3).

Третья формула алгоритма (9) может быть записана как

$$\begin{aligned} \mathbf{P}^{-1}(k|k)[\sigma_v^2 (\hat{\mathbf{S}}(k) - \mathbf{F}\hat{\mathbf{S}}(k-1))] &= \\ &= \mathbf{Z}(k) - \mathbf{F}\hat{\mathbf{S}}(k-1). \end{aligned} \quad (14)$$

Благодаря доказанному свойству матрицы $\mathbf{P}^{-1}(k|k)$, вектор $\hat{\mathbf{S}}(k)$ может быть эффективно определен из формулы (14) путем применения разложения Холецкого к матрице $\mathbf{P}^{-1}(k|k)$.

Итак, разработанный экономичный способ блочной калмановской фильтрации состоит из выполнения на каждом шаге следующих действий:

- обновление матрицы $P^{-1}(k|k)$ в соответствии с формулами (12) и (13);
- нахождение искомой оценки путем решения уравнения (14).

Высокая экономичность данного алгоритма обусловлена отсутствием в нем прямых матричных обращений. Отметим, что полученные результаты могут быть перенесены на случай окрашенного шума, что расширяет сферу их применения.

Рассмотрим вычислительные затраты, связанные с применением предложенного метода для обработки N зашумленных наблюдений. В частном случае, когда размер блока совпадает с периодом квазистационарности речи, требуемое количество элементарных операций составляет порядка p^2N . В то же время, традиционный фильтр Калмана, основанный на модели (3), требует, согласно [1], $4p^2N$ операций. Таким образом, в нашем случае достигается сокращение вычислительных затрат в четыре раза.

Для практического применения описанной процедуры фильтрации необходимо знать значения авторегрессионных параметров речи. В следующем разделе будет рассмотрено оценивание авторегрессионных параметров в присутствии шумов.

3. ОЦЕНИВАНИЕ АВТОРЕГРЕССИОННЫХ ПАРАМЕТРОВ РЕЧИ В ПРИСУТСТВИИ ШУМА

Среди помехоустойчивых методов оценивания авторегрессионных параметров особое распространение получили подходы, основанные на максимизации функционалов правдоподобия и апостериорной вероятности. Существующие методы решения этой задачи имеют, как правило, итерационную структуру и не обеспечивают глобальной максимизации. Их эффективность сильно зависит от выбора начального приближения. Поэтому в работе [4] был предложен новый помехоустойчивый метод оценивания авторегрессионных параметров речи. Суть данного подхода состояла в максимизации функционала правдоподобия, зависящего от характеристик помехи, на ограниченном количестве авторегрессионных квантов $\{\mathbf{a}^{(m)}, g^{(m)}\}$, $m = 1, \dots, M$, вместо максимизации на непрерывном пространстве $\{\mathbf{a}, g\}$ ($\mathbf{a} = [a_1, \dots, a_p]$). Вычисление таких квантов может быть осуществлено с помощью итерационного кластерного алгоритма К-средних [5].

В отличие от традиционных подходов, указанный метод содержит систематическую ошибку

оценивания (смещение). Заметим, однако, что величина смещения определяется такими факторами, как количество квантов и число дикторов, принимавших участие в формировании кодовой книги, и всегда может быть сведена к желаемому уровню. В то же время, присутствие смещения компенсируется снижением статистической ошибки оценивания (дисперсии оценки).

Основное препятствие для применения данной методологии состоит в том, что подсчет функционала правдоподобия для каждого кванта ведет к существенным вычислительным затратам, когда количество квантов относительно велико. В связи с этим целесообразно рассмотреть следующий экономичный способ вычисления функционала правдоподобия, основанный на блочной модели (7).

Предположим, что мы имеем блок \mathbf{Z} из l измерений зашумленного авторегрессионного сигнала. Из формулы (7) следует, что ковариационная матрица вектора \mathbf{Z} равна $g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T} + \sigma_v^2 \mathbf{E}_l$. Поэтому, предполагая, что полезный сигнал и помеха имеют нормальное распределение, функционал правдоподобия представим как

$$p(\mathbf{Z}/\mathbf{a}, g) = (2\pi)^{-0.5l} \times \sqrt{\det(g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T} + \sigma_v^2 \mathbf{E}_l)} \times \exp[-0.5 \mathbf{Z}^T (g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T} + \sigma_v^2 \mathbf{E}_l)^{-1} \mathbf{Z}]. \quad (15)$$

Преобразуем выражение в показателе экспоненты:

$$\begin{aligned} \mathbf{Z}^T (g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T} + \sigma_v^2 \mathbf{E}_l)^{-1} \mathbf{Z} &= \\ &= \mathbf{Z}^T \mathbf{A}^T (g^2 \mathbf{E}_l + \sigma_v^2 \mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{Z}. \end{aligned} \quad (16)$$

Принимая во внимание разложение Холецкого

$$g^2 \mathbf{E}_l + \sigma_v^2 \mathbf{A} \mathbf{A}^T = \mathbf{C}^T \mathbf{C}, \quad (17)$$

запишем формулу (16) в виде

$$\begin{aligned} \mathbf{Z}^T (g^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T} + \sigma_v^2 \mathbf{E}_l)^{-1} \mathbf{Z} &= \\ &= \mathbf{Z}^T \mathbf{A}^T (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{A} \mathbf{Z} = \\ &= (\mathbf{C}^{T,-1} \mathbf{A} \mathbf{Z})^T \mathbf{C}^{T,-1} \mathbf{A} \mathbf{Z}. \end{aligned} \quad (18)$$

Матрица $g^2 \mathbf{E}_l + \sigma_v^2 \mathbf{A}^{-1} \mathbf{A}^{-1,T}$ имеет $(2p+1)$ -диагональную структуру. Поэтому матрица \mathbf{C}^T в разложении (17) является нижней $(p+1)$ -диагональной. Это существенно упрощает вычисление вектора $\mathbf{X} = \mathbf{C}^{T,-1} \mathbf{A} \mathbf{Z}$, эквивалентное решению уравнения $\mathbf{C}^T \mathbf{X} = \mathbf{A} \mathbf{Z}$. Легко видеть, что выражение (18) – сумма квадратов координат

вектора \mathbf{X} . Наконец, из формулы (17) следует, что

$$\sqrt{\det(g^2 \mathbf{A}^{-1} \mathbf{A}^{-1.T} + \sigma_v^2 \mathbf{E}_l)} = \sqrt{\det \mathbf{C}^T \mathbf{C}} = \prod_{k=1}^l C_{kk}.$$

Указанной процедурой завершается вычисление функционала (15).

Другой возможный способ сокращения вычислительных затрат состоит в применении комбинированного двухэтапного алгоритма. На первом этапе результат прямой максимизации функционала правдоподобия на небольшом количестве квантов используется лишь в качестве начального приближения. Точное же значение оценки определяется с помощью хорошо известной итерационной процедуры [2], обеспечивающей сходимость к глобальному максимуму функционала правдоподобия при условии хорошей инициализации. На каждой итерации этого алгоритма выполняется решение модифицированных уравнений Юла–Уокера с последующим оптимальным линейным оцениванием сигнала на основе полученных оценок параметров.

Основной недостаток указанной итерационной процедуры состоит в ее высоких вычислительных затратах, связанных с вычислением оптимальной линейной оценки сигнала на каждой итерации. Эту оценку традиционно получают путем двукратного применения фильтра Калмана (в прямом и обратном направлениях) [2], что требует выполнения очень большого количества операций. Заметим, что последнее препятствие может быть полностью преодолено с помощью экономичного метода фильтрации, описанного в разделе 2.

Описанный комбинированный подход позволяет снизить общую погрешность оценивания по сравнению с прямой максимизацией функционала правдоподобия, благодаря исключению систематической компоненты ошибки. Кроме того, данный алгоритм решает важную проблему итерационных подходов, связанную с выбором начального приближения. В следующем разделе будет показано, что количество квантов, необходимое для качественной инициализации, может быть очень небольшим.

4. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Как было показано в разделе 2, предложенная процедура блочной фильтрации обеспечивает сокращение вычислений примерно в четыре раза, по сравнению с традиционным фильтром Калмана. Следующий шаг состоит в проверке эффективно-

сти разработанного метода коррекции при работе с реальными речевыми сигналами.

В экспериментах использовались русскоязычные фонетически сбалансированные предложения, произносимые четырьмя дикторами-мужчинами и двумя дикторами-женщинами, дискретизированные с частотой 8000 Гц. Эти сигналы были смешаны с белым шумом, соответствующим различным отношениям сигнал/помеха (ОСП). Оценивание авторегрессионных параметров производилось на неперекрывающихся фреймах длиной 20 мс ($L=160$ дискретных отсчетов). Порядок авторегрессионной модели p был выбран равным десяти.

Для начала рассмотрим случай, когда параметры сигнала и шума предполагаются известными (авторегрессионные параметры речи были определены на каждом фрейме заранее с помощью автокорреляционного метода [6]). В табл. 1 приводятся итоговые ОСП, полученные при обработке зашумленного сигнала следующими методами:

- традиционный фильтр Калмана (ТФК), основанный на модели (3);
- блочный фильтр Калмана (БФК), соответствующий длинам блоков $l=10=p$, $l=40$, $l=160=L$.

Из табл. 1 видно, что модификации блочно-го фильтра Калмана имеют существенное преимущество перед традиционным алгоритмом (от $1.0 \div 1.3$ дБ при ОСП=20 дБ до $1.7 \div 2.3$ дБ при ОСП=0 дБ). Также видно, что повышение длины блока приводит к снижению ошибки, поскольку в процедуру фильтрации вовлекается большее число наблюдений.

Перейдем к случаю, когда параметры сигнала и шума изначально неизвестны. Для оценивания авторегрессионных параметров речи применялась двухэтапная процедура, описанная в разделе 3. На первом этапе использовались кодовые книги различных размеров (4, 16, 64, 128), сгенерированные с помощью дикторов, которые не принимали участия в формировании тестовых сигналов. На втором этапе для итерационного улучшения предварительных оценок использовалось пять итераций. Оценивание дисперсии шума производилось на начальных 100 мс, не содержащих речевого сигнала.

В табл. 2 показаны соответствующие итоговые ОСП, полученные при обработке зашумленного сигнала блочным фильтром Калмана с длиной блока $l=L=160$, при различных входных ОСП. Для каждого количества квантов приводятся ОСП, соответствующие первому (в скобках) и

Табл. 1. Итоговые ОСП для исследуемых методов фильтрации (случай изначально известных авторегрессионных параметров)

Входное ОСП	ТФК	БФК(10)	БФК(40)	БФК(160)
0 дБ	7.78	9.46	9.97	10.12
5 дБ	10.96	12.80	13.26	13.40
10 дБ	14.42	16.06	16.47	16.63
15 дБ	18.22	19.59	19.93	19.98
20 дБ	22.23	23.27	23.44	23.52

Табл. 2. Итоговые ОСП для различных модификаций предложенного метода оценивания авторегрессионных параметров

Входное ОСП	4 кванта	16 квантов	64 кванта	128 квантов
0 дБ	9.22 (8.52)	9.38 (9.09)	9.31(9.17)	9.32 (9.19)
5 дБ	12.44 (11.55)	12.62 (12.23)	12.67 (12.47)	12.64 (12.49)
10 дБ	15.77 (14.86)	16.02 (15.52)	16.09 (15.73)	16.09 (15.80)
15 дБ	19.48 (18.41)	19.64 (18.95)	19.70 (19.21)	19.70 (19.27)

второму этапам оценивания.

Как видно из табл. 2, для инициализации итерационного алгоритма вполне достаточно брать 16 квантов. Последующее увеличение количества квантов практически не влияет на величину ошибки фильтрации. Следует отметить, что все значения выходных ОСП отличаются от значений, полученных в “идеальном” случае (при известных параметрах), менее, чем на 1 дБ. Это свидетельствует об эффективности предложенного метода оценивания параметров и метода очистки речи в целом.

ЗАКЛЮЧЕНИЕ

Предложен экономичный метод очистки речи от шума, основанный на блочной модели авторегрессионного сигнала в пространстве состояний. Достигнуто сокращение вычислительных затрат в четыре раза, по сравнению с традиционным фильтром Калмана. Показано уменьшение ошибки фильтрации. Разработан эффективный двухэтапный алгоритм оценивания авторегрессионных параметров речи в присутствии шума. Эффектив-

ность результирующего метода очистки проверена на реальных речевых сигналах.

1. Goh Z., Tan K.-C., Tan B. Kalman filtering speech enhancement method based on voiced/unvoiced speech model // IEEE Trans. Speech Audio Proces.– 1999.– 7.– P. 510–525.
2. Gannot S., Burnstein D., Weinstein E. Iterative and sequential Kalman filter-based speech enhancement algorithms // IEEE Trans. Speech Audio Proces.– 1998.– 6.– P. 373–385.
3. Lim J., Oppenheim A. All-pole modeling of degraded speech // IEEE Trans. Acoust. Speech Signal Proces.– 1978.– 26.– P. 197–210.
4. Kovtonyuk A., Kalyuzhny A., Semenov V. Adaptive Kalman filtering of speech signals based on a block model in the state space and vector quantization of autoregressive features // Proc. 8-th Austral. Conf. SST-2000.– Canberra, 2000.– P. 262–267.
5. Макхоул Дж., Русос С., Гиш Г. Векторное квантование при кодировании речи // ТИИЭР.– 1985.– 73.– С. 19–61.
6. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов.– М.: Радио и связь, 1981.– 365–381 с.
7. Сейдж А., Мэлс Дж. Теория оценивания и ее применение в связи и управлении.– М.: Радио и связь, 1976.– С. 251–317.
8. Хорн Р. А., Джонсон С. Р. Матричный анализ.– М.: Мир, 1989.– 655 с.