

УДК 534.78+621.391

МЕТОД ИДЕНТИФИКАЦИИ АВТОРЕГРЕССИОННЫХ ПАРАМЕТРОВ РЕЧИ В ЧАСТОТНОЙ ОБЛАСТИ

А. Я. КАЛЮЖНЫЙ, А. А. КОВТОНЮК, В. Ю. СЕМЕНОВ

Государственное научно-производственное предприятие “Дельта”, Киев

Получено 18.05.2010

Рассмотрена задача оценивания авторегрессионных параметров зашумленного речевого сигнала в частотной области. На основании частотного представления авторегрессионного сигнала проиллюстрированы методика расчета функции правдоподобия, а также реализация метода ожидания-максимизации для итеративного оценивания авторегрессионных параметров. Анализ различных мер искажения речевых сигналов показал, что работа в частотной области обеспечивает такую же точность, как и соответствующие подходы во временной области, но при существенно меньших вычислительных затратах. Предложен и исследован двухэтапный метод оценивания авторегрессионных параметров зашумленного речевого сигнала в частотной области.

Розглянуто задачу оцінювання авторегресійних параметрів зашумленого мовного сигналу в частотній області. На базі частотного представлення авторегресійного сигналу проілюстровано методику розрахунку функції правдоподібності, а також реалізацію методу очікування-максимізації для ітеративного оцінювання авторегресійних параметрів. Аналіз різних мір спотворення мовних сигналів показує, що робота в частотній області забезпечує таку саму точність, як і відповідні підходи в часовій області, але при істотно менших обчислювальних затратах. Окрім того, запропоновано й досліджено двоетапний метод оцінювання авторегресійних параметрів зашумленого мовного сигналу в частотній області.

The paper deals with considering of estimation of noisy speech autoregressive parameters in frequency domain. The method for calculating of likelihood function and implementation of the Expectation-Maximization method for the iterative estimation of speech autoregressive parameters have been illustrated on the base of frequency representation of the autoregressive signal. The analysis of different speech distortion measures shows that operation in frequency domain provides the similar accuracy as the corresponding approaches in time domain, but with essentially lower computational expenses. Moreover, a two-stage method for estimating of speech autoregressive parameters in frequency domain has been presented and investigated.

ВВЕДЕНИЕ

Авторегрессионные модели широко используются при обработке аудиосигналов и изображений, анализе экономических процессов и в других областях, в частности, при цифровой обработке речевых сигналов. Большинство современных методов речевой обработки основаны на авторегрессионной модели речеобразования, которая имеет следующее представление во временной области:

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

где $s(n)$ – дискретизированный речевой сигнал; $w(n)$ – возбуждающий процесс; g – коэффициент усиления; a_k ($k = 1, \dots, p$) – авторегрессионные коэффициенты, определяющие спектральную огибающую речевого сигнала [1].

Известно, что эффективность систем обработки речи, основанных на вычислении авторегрессионных параметров, резко ухудшается в присутствии фонового шума, т. е. когда доступен для измерения не сам речевой сигнал $s(n)$, а его сумма с шумом $v(n)$:

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

Это обусловило большую актуальность разработки помехоустойчивых методов оценивания авторегрессионных параметров речи. Этой проблеме посвящено большое количество исследовательских работ [2–8].

Большинство помехоустойчивых методов оценивания авторегрессионных параметров являются разновидностями метода ожидания-максимизации (Expectation-Maximization – EM), позволяющего получить оценку максимального правдоподобия [9]. Точная реализация этого метода применительно к оцениванию авторегрессионных параметров речи, которая может считаться эталоном для сравнения с остальными алгоритмами, приведена в работе [5]. Общая проблема итеративного подхода метода EM заключается в необходимости иметь хорошее начальное приближение, так как для низких отношений сигнал-шум некачественная инициализация метода EM может дать сходимость к локальному минимуму вместо локального максимума [5]. Заметим также, что эталонный метод характеризовался высокими вычислительными затратами калмановского сглаживания, поскольку все расчеты выполнялись во временной области.

В работах [6–8] нами предложен двухэтапный

метод оценивания авторегрессионных параметров, в котором начальное приближение для метода ЕМ строится путем максимизации функции правдоподобия на ограниченном множестве авторегрессионных квантов, предварительно вычисленных с помощью алгоритма K -средних. Однако сопутствующие вычислительные затраты оказываются достаточно высокими (более 100 миллионов операций с плавающей точкой) и могут существенно возрасти при реализации на цифровых сигнальных процессорах с фиксированной точкой.

Учитывая сказанное, в данном исследовании мы предлагаем видоизменить методику оценивания авторегрессионных параметров путем замены модели авторегрессионного сигнала во временной области моделью в частотной области. Как будет показано, это приводит к существенному сокращению вычислительных затрат по расчету функции правдоподобия и реализации метода ЕМ. При этом предложенные методы сохраняют такую же точность вычислений, что и методы, работающие во временной области.

В первом и втором разделах статьи рассмотрена блочная модель авторегрессионного сигнала в частотной области и расчет функции правдоподобия на основании этой модели. Третий раздел посвящен реализации метода ЕМ в частотной области. В четвертом разделе производится анализ вычислительных затрат предложенных методов и формулируется двухэтапный метод оценивания авторегрессионных параметров, работающий в частотной области. В экспериментальном разделе работа метода оценивается по различным критериям качества обработки речевых сигналов.

1. БЛОЧНАЯ МОДЕЛЬ АВТОРЕГРЕССИОННОГО СИГНАЛА

Для реализации метода оценивания авторегрессионных параметров в работах [6–8] рассматривалась блочная модель зашумленного авторегрессионного сигнала (1), (2) во временной области:

$$\mathbf{Z}(k) = g\mathbf{A}^{-1}\mathbf{W}(k) + \mathbf{V}(k), \quad (3)$$

где $\mathbf{S}(k), \mathbf{W}(k), \mathbf{V}(k), \mathbf{Z}(k)$ – соответственно блоки из L значений речевого сигнала, возбуждающего процесса, шума и наблюдений на k -ом фрейме,

$$\begin{aligned} \mathbf{S}(k) &= [s((k-1)L+1), s((k-1)L+2), \dots, s(kL)]^T, \\ \mathbf{W}(k) &= [w((k-1)L+1), w((k-1)L+2), \dots, w(kL)]^T, \\ \mathbf{V}(k) &= [v((k-1)L+1), v((k-1)L+2), \dots, v(kL)]^T, \\ \mathbf{Z}(k) &= [z((k-1)L+1), z((k-1)L+2), \dots, z(kL)]^T. \end{aligned}$$

Отбеливающая теплицева матрица заполнена исходными авторегрессионными коэффициентами модели (1):

$$\mathbf{A}_{ij} = \begin{cases} 1, & i = j, \\ a_{i-j}, & j < i \leq j + p, \\ 0, & i < j \text{ или } i > j + p. \end{cases} \quad (4)$$

Предполагается, что помеха окрашена и характеризуется собственными авторегрессионными коэффициентами b_1, \dots, b_q , т. е. вектор \mathbf{V} может быть представлен как

$$\mathbf{V}(k) = h\mathbf{B}^{-1}\mathbf{U}(k), \quad (5)$$

где усиление h и отбеливающая матрица \mathbf{B} имеют тот же смысл, что и g и \mathbf{A} для речевого сигнала.

В данном исследовании мы предлагаем заменить модель (3) представлением в частотной области:

$$\mathbf{Z}_\omega(n) = \frac{g\mathbf{W}_\omega(n)}{\mathbf{A}_\omega(n)} + \frac{h\mathbf{U}_\omega(n)}{\mathbf{B}_\omega(n)}, \quad (6)$$

где нижние индексы “ ω ” означают преобразования Фурье исходных массивов:

$$\begin{aligned} \mathbf{Z}_\omega(n) &= \sum_{l=1}^L z(l)e^{-2\pi iln/L}, & n = 1, \dots, L; \\ \mathbf{W}_\omega(n) &= \sum_{l=1}^L w(l)e^{-2\pi iln/L}, & n = 1, \dots, L; \\ \mathbf{U}_\omega(n) &= \sum_{l=1}^L u(l)e^{-2\pi iln/L}, & n = 1, \dots, L. \end{aligned}$$

Матрицы $\mathbf{A}_\omega(n)$ и $\mathbf{B}_\omega(n)$ определяются аналогичным образом:

$$\begin{aligned} \mathbf{A}_\omega(n) &= 1 + \sum_{k=1}^p a_k e^{-2\pi i kn/L}, & n = 1, \dots, L; \\ \mathbf{B}_\omega(n) &= 1 + \sum_{k=1}^q b_k e^{-2\pi i kn/L}, & n = 1, \dots, L. \end{aligned}$$

Фактически, представление в частотной области (6) эквивалентно замене в модели (3) теплицевых матриц \mathbf{A} и \mathbf{B} на соответствующие им циркулянтные матрицы [10].

2. РАСЧЕТ ФУНКЦИИ ПРАВДОПОДОБИЯ В ЧАСТОТНОЙ ОБЛАСТИ

На основе модели (6) функция правдоподобия авторегрессионных параметров записывается в ви-

де

$$p(\mathbf{Z}_\omega/\theta) = (2\pi)^{-0.5L} [\det \text{Cov}\{\mathbf{Z}_\omega, \mathbf{Z}_\omega\}]^{-0.5} \times \exp[-0.5 \mathbf{Z}_\omega^T \text{Cov}^{-1}\{\mathbf{Z}_\omega, \mathbf{Z}_\omega\} \mathbf{Z}_\omega], \quad (7)$$

где вектор θ включает в себя авторегрессионные параметры сигнала и шума,

$$\theta = [a_1, \dots, a_p, g, b_1, \dots, b_q, h].$$

Из соотношения (6) получим выражение для ковариационной матрицы наблюдений в частотной области

$$\text{Cov}\{\mathbf{Z}_\omega, \mathbf{Z}_\omega\} = L \text{diag}\{P_s + P_v\}, \quad (8)$$

где P_s – квадрат спектра (амплитудно-частотной характеристики) авторегрессионного фильтра с коэффициентами a_k ($k=1, \dots, p$) и усилением g ,

$$P_s(n) = \frac{g^2}{|\mathbf{A}_\omega(n)|^2}, \quad n = 1, \dots, L. \quad (9)$$

Аналогично определяется массив P_v :

$$P_v(n) = \frac{h^2}{|\mathbf{B}_\omega(n)|^2}, \quad n = 1, \dots, L. \quad (10)$$

В отличие от представления во временной области, ковариационная матрица (8) является диагональной, что упрощает последующие расчеты.

Подставляя выражения (8)–(10) в формулу (7), получаем

$$p(\mathbf{Z}_\omega/\theta) = (2\pi)^{-0.5L} \times \left[\prod_{n=1}^L \delta_n \right]^{-0.5} \exp \left[-0.5 \sum_{n=1}^L \delta_n |\mathbf{Z}_\omega(n)|^2 \right], \quad (11)$$

где весовые коэффициенты δ_n ($n=1, \dots, L$) определяются следующим образом:

$$\delta_n = \frac{1}{L(P_s(n) + P_v(n))}.$$

Более удобен для записи логарифм функции правдоподобия:

$$\log p(\mathbf{Z}_\omega/\mathbf{a}, g) = C - 0.5 \sum_{n=1}^L \log \delta_n - 0.5 \sum_{n=1}^L \delta_n |\mathbf{Z}_\omega(n)|^2, \quad (12)$$

где C – некоторая константа, не оказывающая влияния на его максимизацию.

3. РЕАЛИЗАЦИЯ МЕТОДА ЕМ В ЧАСТОТНОЙ ОБЛАСТИ

С формальной точки зрения метод ЕМ требует определения так называемых “неполных” и “полных” данных [9]. Пусть вектор \mathbf{Z}_ω представляет “неполные”, а вектор \mathbf{X} – “полные” данные:

$$\mathbf{X} = [\mathbf{S}_\omega, \mathbf{V}_\omega]. \quad (13)$$

На Е-шаге метода ЕМ вычисляется апостериорное значение логарифма функции правдоподобия вектора полных данных:

$$Q(\theta, \hat{\theta}^{(l)}) = E \left\{ \log p(\mathbf{X}/\theta) / \mathbf{Z}_\omega, \hat{\theta}^{(l-1)} \right\}, \quad (14)$$

где l – номер итерации метода ЕМ.

Выражение для $\log p(\mathbf{X}/\theta)$ может быть получено по аналогии с формулой (12). Согласно [11, (6.2)] в предположении об устойчивости авторегрессионных моделей сигнала и шума, используя соотношения

$$\sum_{n=1}^L \log |\mathbf{A}_\omega(n)| \approx 0, \\ \sum_{n=1}^L \log |\mathbf{B}_\omega(n)| \approx 0$$

и опуская не зависящие от θ константы, получаем

$$\log p(\mathbf{X}/\theta) = -L \log g - 0.5g^{-2} \sum_{n=1}^L |\mathbf{S}_\omega(n)|^2 |\mathbf{A}_\omega(n)|^2 - \\ -L \log h - 0.5h^{-2} \sum_{n=1}^L |\mathbf{V}_\omega(n)|^2 |\mathbf{B}_\omega(n)|^2. \quad (15)$$

При выводе этой формулы учитывалось свойство некоррелированности сигнала и шума.

Подставляя соотношения (15) в (14), получаем

$$E \{ \log p(\mathbf{X}/\theta) / \mathbf{Z}_\omega \} = -L \log g - 0.5g^{-2} \times \\ \times \sum_{n=1}^L E \{ |\mathbf{S}_\omega(n)|^2 / \mathbf{Z}_\omega \} |\mathbf{A}_\omega(n)|^2 - \\ -L \log h - 0.5h^{-2} \times \\ \times \sum_{n=1}^L E \{ |\mathbf{V}_\omega(n)|^2 / \mathbf{Z}_\omega \} |\mathbf{B}_\omega(n)|^2. \quad (16)$$

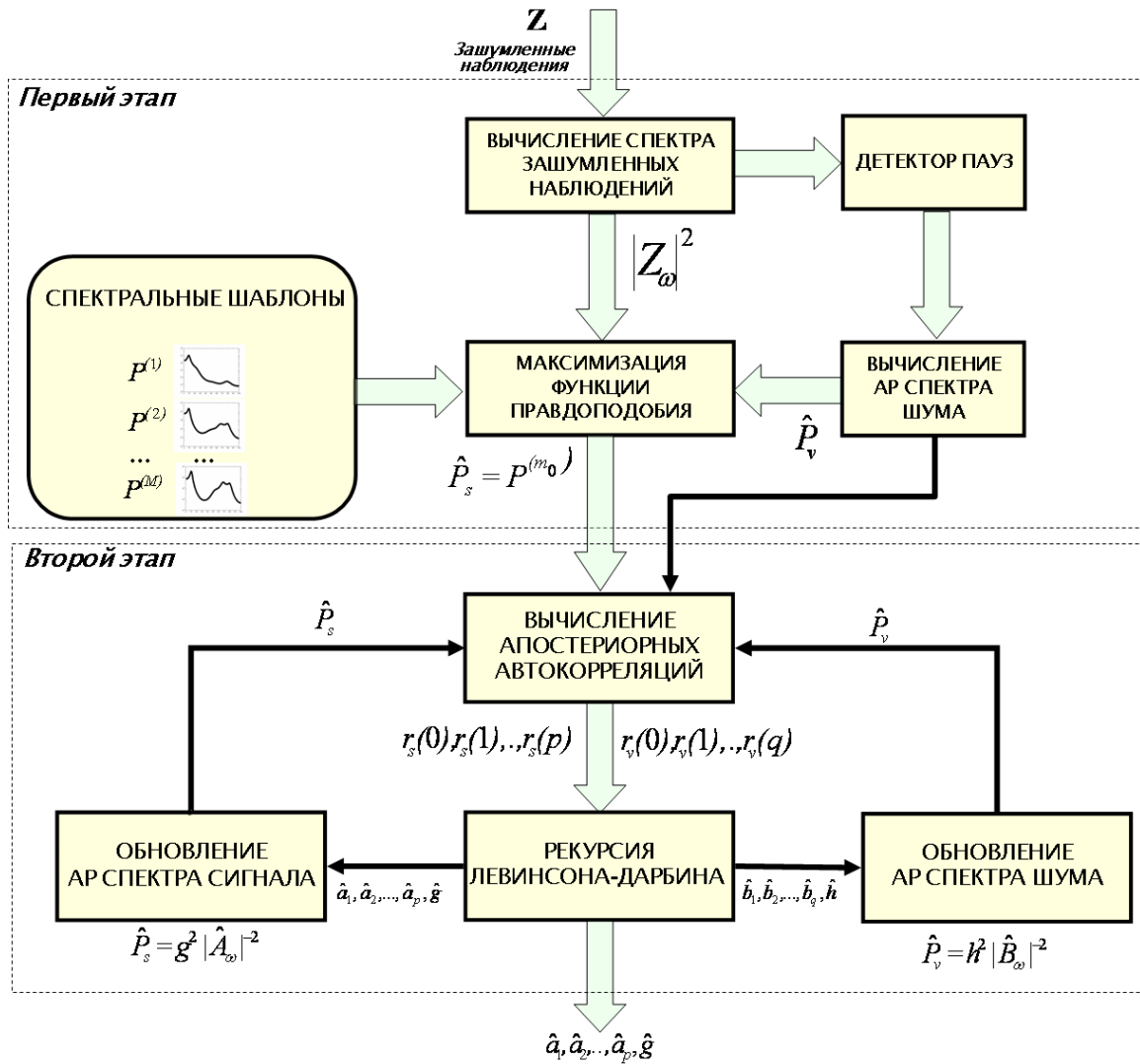


Рисунок. Блок-схема предложенного метода

Величины $E\{|\mathbf{S}_\omega(n)|^2/\mathbf{Z}_\omega\}$, $E\{|\mathbf{V}_\omega(n)|^2/\mathbf{Z}_\omega\}$ могут быть получены по формуле для апостериорной корреляции [14]:

$$E\{|\mathbf{S}_\omega(n)|^2/\mathbf{Z}_\omega\} = L \frac{P_s P_v}{(P_s + P_v)} + |\mathbf{Z}_\omega|^2 \frac{P_s^2}{(P_s + P_v)^2}, \quad (17)$$

$$E\{|\mathbf{V}_\omega(n)|^2/\mathbf{Z}_\omega\} = L \frac{P_s P_v}{(P_s + P_v)} + |\mathbf{Z}_\omega|^2 \frac{P_v^2}{(P_s + P_v)^2}. \quad (18)$$

При этом спектры P_s , P_v определяются согласно формул (9), (10) через оценки параметров с предыдущей итерации.

На M -шаге метода EM максимизируется выражение (16). Дифференцирование (16) по a_1, \dots, a_p, g приводит к следующим соотношениям для оценок авторегрессионных параметров:

$$\sum_{j=1}^p r_s(i-j)\hat{a}_j = -r_s(i), \quad i = 1, \dots, p, \quad (19)$$

$$\hat{g} = \sqrt{\frac{1}{L} \left[r_s(0) + \sum_{k=1}^p \hat{a}_k r_s(k) \right]}, \quad (20)$$

где коэффициенты $r_s(k)$ ($k=0, \dots, p$) – первые $p+1$ элементов обратного преобразования Фурье от массива (17).

Аналогично, формулы для оценок авторегресси-

онных параметров шума имеют вид

$$\sum_{j=1}^q r_v(i-j)\hat{b}_j = -r_v(i), \quad i = 1, \dots, q, \quad (21)$$

$$\hat{h} = \sqrt{\frac{1}{L} \left[r_v(0) + \sum_{k=1}^q \hat{b}_k r_v(k) \right]}, \quad (22)$$

где коэффициенты $r_v(k)$, ($k=0, \dots, q$) – первые $q+1$ элементов обратного преобразования Фурье от массива (18).

Системы (19) и (21) являются теплицевыми и могут быть решены рекурсивным методом Левинсона – Дарбина [1].

Таким образом, повторяя итеративно E- и M-шаги, можно получить искомые оценки авторегрессионных параметров речевого сигнала $\hat{a}_1, \dots, \hat{a}_p$, \hat{g} и шума $\hat{b}_1, \dots, \hat{b}_q$, \hat{h} .

4. СТРУКТУРА ПРЕДЛАГАЕМОГО МЕТОДА ОЦЕНИВАНИЯ АВТОРЕГРЕССИОННЫХ ПАРАМЕТРОВ И АНАЛИЗ ВЫЧИСЛИТЕЛЬНЫХ ЗАТРАТ

Новая методика вычисления функции правдоподобия и реализации алгоритма EM позволяет изменить процедуру оценивания авторегрессионных параметров, изложенную в [7, 8]. Общая структура предлагаемого двухэтапного метода показана на рисунке.

Как видно из блок-схемы, все вычисления проводятся со спектрами мощности сигналов и их авторегрессионных моделей. Так, на первом этапе оперируют спектрами квантов авторегрессионных параметров и спектром наблюдений. Фактически, алгоритм проверяет насколько спектр каждого из квантов аппроксимирует спектр наблюдений по критерию максимального правдоподобия. Заметим, что спектры авторегрессионных квантов могут быть вычислены заранее и храниться в постоянном запоминающем устройстве. На втором этапе проводится итеративное уточнение спектра авторегрессионной модели. Начальные оценки для авторегрессионных параметров шума могут быть получены с помощью детектора пауз и последующего усреднения их спектра мощности (см., например, [12]).

Поскольку основными операциями метода являются быстрые преобразования Фурье, общее их количество существенно зависит от числа машинных циклов, требуемых для одного такого преобразования. В дальнейшем предполагаем, что для обработки фрейма длиной L (если L – степень двух)

последнее составляет

$$N_{\text{БПФ}} = \alpha L \log_2 L,$$

где α – коэффициент, зависящий от типа вычислительной платформы. Мы рассмотрели три его возможных величины:

- $\alpha = 3$ – теоретическое значение [10];
- $\alpha = 1$;
- $\alpha = 0.3$ – значение, достижимое благодаря аппаратной реализации быстрого преобразования Фурье на цифровых сигнальных процессорах.

Общее количество операций в секунду для предложенного метода оценивания в частотной области составляет

$$N = N_1 + N_2 = f_s \left[\alpha \log_2 L + 4M + K \left(14 + 4\alpha \log_2 L + \frac{4p^2 + 4q^2 + 3p + 3q}{L} \right) \right], \quad (23)$$

где N_1 , N_2 – количество операций для первого и второго этапов метода во временной области соответственно; f_s – частота дискретизации; M – число квантов; K – число итераций. Это означает, что для базовой конфигурации $M = 16$ квантов, $K = 5$ итераций EM, частоты дискретизации $f_s = 8000$ Гц, порядков авторегрессионных моделей $p = 10$, $q = 8$, длины фрейма $L = 256$ (32 мс) и $\alpha = 1$ величина (23) составляет примерно 2.5 миллиона операций в секунду.

В то же время, затраты двухэтапного метода, работающего во временной области [7, 8], составляли порядка

$$N' = N'_1 + N'_2 = f_s \left[M(p^2 + 8p) + K \left(8p^2 + 8pq + \frac{9q^2}{2} + \frac{p^3}{3} + \frac{41p}{3} + \frac{29q}{2} \right) \right], \quad (24)$$

где N'_1 , N'_2 – количество операций для первого и второго этапов метода в частотной области.

Сравнение вычислительных затрат методов во временной и в частотной областях представлено в табл. 1. Из нее видно, что работа в частотной области позволяет сократить число операций в (22.3...73.2) раз для различных α . Что касается затрат собственно метода EM, они сокращаются соответственно в (20.5...87.7) раз.

Табл. 1. Итоговая экономия вычислительных затрат

Относительное число операций	α		
	0.3	1	3
N'_1/N_1	44.1	40.9	33.5
N'_2/N_2	87.7	47.4	20.5
N'/N	73.2	45.9	22.3
$N'/(N + N_{\text{БФК}})$	33.2	26.2	16.3

Табл. 2. Сравнение мер спектрального искажения для различных методов

Разновидность метода ЕМ	Отношение сигнал-шум, дБ			
	5	10	15	20
[7, 8] (ВО)	3.77	3.10	2.50	2.00
Предлагаемая (ЧО)	3.79	3.10	2.47	1.90
[5] (ВО)	4.24	3.34	2.55	1.91

Табл. 3. Сравнение мер Итакуры – Сайто для различных методов

Разновидность метода ЕМ	Отношение сигнал-шум, дБ			
	5	10	15	20
[7, 8] (ВО)	0.79	0.48	0.33	0.24
Предлагаемая (ЧО)	0.78	0.48	0.30	0.20
[5] (ВО)	1.11	0.66	0.39	0.22

Табл. 4. Сравнение нормированных мер Итакуры – Сайто для различных методов

Разновидность метода ЕМ	Отношение сигнал-шум, дБ			
	5	10	15	20
[7, 8] (ВО)	0.36	0.26	0.18	0.12
Предлагаемая (ЧО)	0.37	0.26	0.18	0.11
[5] (ВО)	0.46	0.29	0.18	0.10

Табл. 5. Сравнение сегментных отношений сигнал-шум для различных методов

Разновидность метода ЕМ	Отношение сигнал-шум, дБ			
	5	10	15	20
[7, 8] (ВО)	5.42	8.53	12.05	15.99
Предлагаемая (ЧО)	5.36	8.45	12.00	15.96
Предлагаемая (ЧО) + БФК	5.36	8.54	12.13	16.10
[5] (ВО)	4.20	7.97	11.9	16.06

Если требуется не только оценить авторегрессионные параметры речи, но и восстановить исходный речевой сигнал, его оценка может быть получена с помощью блочного фильтра Калмана [7, 8], что приводит к выполнению

$$N_{\text{БФК}} = ((p+1)(p+8) + 5q)f_s \quad (25)$$

дополнительных операций (см. последнюю строку табл. 1).

5. АНАЛИЗ ТОЧНОСТИ ПРЕДЛОЖЕННОГО ПОДХОДА

Проверим, приводят ли методы, основанные на временной (3) и частотной (6) моделях авторегрессионного сигнала, к одинаковым оценкам авторегрессионных параметров. В экспериментах использовались русскоязычные сбалансированные предложения, произносимые четырьмя дикторами-мужчинами и двумя дикторами-женщинами и оцифрованные с частотой 8000 Гц. Эти сигналы были смешаны с аддитивной окрашенной помехой при различных отношениях сигнал-шум – 5, 10, 15 и 20 дБ. Оценивание авторегрессионных параметров проводилось на фреймах продолжительностью 32 мс ($L=256$ отсчетов) для порядков авторегрессионных моделей сигнала и шума $p=10$ и $q=8$ соответственно.

Были проанализированы следующие методы:

- 1) метод EM во временной области (ВО) с инициализацией, предложенной в [7, 8];
- 2) предлагаемый метод, основанный на модели в частотной области (ЧО);
- 3) метод EM во временной области (ВО) [5] с инициализацией по методу линейного предсказания.

Для двух первых методов рассматривалась конфигурация с $M=16$ квантов. Для реализации метода EM использовались $K=5$ его итераций. Методы сопоставлялись на основании следующих критериев:

- 1) среднее спектральное искажение [13];
- 2) мера Итакуры–Саито [15];
- 3) нормированная мера Итакуры–Саито (логарифм отношения правдоподобия) [15];
- 4) сегментное отношение сигнал-шум при очистке речи от шума с помощью блочного фильтра Калмана [7, 8], основанного на полученных оценках авторегрессионных параметров.

Как видно из табл. 2–5, при существенной экономии вычислительных ресурсов предложенный метод по всем проанализированным критериям обеспечивает практически такую же точность, что и метод, работающий во временной области [7, 8].

ВЫВОДЫ

1. Предложен подход к оцениванию параметров авторегрессионной модели речевого сигнала, основанный на ее представлении в частотной области.
2. На основании частотной модели разработана методика расчета функции правдоподобия авторегрессионных параметров. Также изложена реализация алгоритма ожидания-максимизации (Expectation-Maximization) для оценивания авторегрессионных параметров в частотной области.
3. Анализ различных мер искажения речевых сигналов показал, что предложенные подходы в частотной области обладают такой же точностью, что и соответствующие им подходы во временной области, но характеризуются существенно меньшими вычислительными затратами.
4. Сформулирован и исследован двухэтапный метод оценивания авторегрессионных параметров зашумленного речевого сигнала, являющийся частотной версией ранее предложенного авторами метода, оперирующего с временной областью.

1. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 496 с.
2. Lim J. S., Oppenheim A. V. All-pole modeling of degraded speech // IEEE Trans. Acoust. Speech Signal Audio Proces. – 1978. – 26. – P. 197–210.
3. Sreenivas T. V., Kirnapure P. Codebook constrained Wiener filtering for speech enhancement // IEEE Trans. Speech Audio Audio Proces. – 1996. – 4. – P. 383–389.
4. Pellom B. L., Hansen J. H. L. An improved (Auto:I, LSP:T) constrained iterative speech enhancement for colored noise environments // IEEE Trans. Speech Audio Audio Proces. – 1998. – 6. – P. 573–579.
5. Gannot S., Burnstein D., Weinstein E. Iterative and sequential Kalman filter-based speech enhancement algorithms // IEEE Trans. Speech Audio Audio Proces. – 1998. – 6. – P. 373–385.
6. 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. Int. Conf. Speech Science Technology. – Canberra, 2000. – P. 262–267.

7. *Semenov V., Kovtonyuk A., Kalyuzhny A.* Computationally efficient method of speech enhancement based on block representation of signal in state space and vector quantization // Proc. Int. Conf. Spoken Language Audio Proces.– Denver, 2002.– P. 2509–2512.
8. *Семенов В. Ю.* Разработка адаптивных методов коррекции речевых сигналов на основе авторегрессионной модели голосового тракта. Дис. ... канд. физ.-мат. наук.– К.: Ин-т гидромех. НАНУ, 2004.– 172 с.
9. *Dempster A., Lair N., Rubin D.* Maximum likelihood from incomplete data via the EM algorithm // J. Roy. Statistic. Soc.– 1977.– **39**.– P. 1–38.
10. *Воеводин В. В., Тыртышников Е. Е.* Вычислительные процессы с теплицевыми матрицами.– М.: Наука, 1987.– 320 с.
11. *Маркел Дж. Д., Грей А. Х.* Линейное предсказание речи.– М.: Связь, 1980.– 308 с.
12. *3GPP TS 46.082.* Voice Activity Detector (VAD) for Enhanced Full Rate (EFR) speech traffic channels.– 3GPP: France, 2004.– 18 p.
13. *Paliwal K. K., Atal B. S.* Efficient vector quantization of LPC parameters at 24 bits/frame // IEEE Trans. Speech Audio Audio Proces.– 1993.– **1**.– P. 3–14.
14. *Сейдж А. П., Мелс Дж. Л.* Теория оценивания и ее применение в связи и управлении.– М.: Связь, 1976.– 496 с.
15. *Hansen J. H. L., Pellom B.* An effective quality evaluation protocol for speech enhancement algorithms // Proc. Int. Conf. Spoken Language Audio Proces.– Sydney, 1998.– P. 2819–2822.