НОВІ ЗАСОБИ
КІБЕРНЕТИКИ, ІНФОРМАТИКИ,
ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ
ТА СИСТЕМНОГО АНАЛІЗУ

УДК 534.78, 621.391.266

В.К. ЗАДІРАКА
Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *zvk140@ukr.net.*

В.Ю. СЕМЕНОВ
Київський академічний університет, ТОВ «Дельта СПЕ», Київ, Україна,
e-mail: *vasyl.delta@gmail.com*.

Є.В. СЕМЕНОВА
Інститут математики НАН України, Київський академічний університет, Київ, Україна,
e-mail: *semenovaevgen@gmail.com.*

МЕТОД ЗАВАДОСТІЙКОГО ОЦІНЮВАННЯ ПАРАМЕТРІВ АВТОРЕГРЕСИВНОЇ МОДЕЛІ В ЧАСТОТНІЙ ОБЛАСТІ

Анотація. Розглянуто задачу оцінювання параметрів авторегресивного сигналу за наявності фонового шуму. На основі частотного представлення такого сигналу показано методику обчислення функціоналу правдоподібності авторегресивних параметрів, а також розглянуто реалізацію методу Expectation-Maximization для їхнього ітеративного оцінювання. Аналіз різних мір спотворення мовленнєвих сигналів показав, що запропоновані підходи у частотній області мають однакову точність із відповідними підходами у часовій області, але характеризуються істотно меншими обчислювальними витратами.

Ключові слова: авторегресивна модель, функція правдоподібності, метод Expectation-Maximization, швидке перетворення Фур'є.

ВСТУП

Розвиток сучасних засобів цифрових комунікацій, таких як стільниковий зв'язок, супутникова та IP-телефонія, зумовлює необхідність розроблення високоефективних алгоритмів кодування мовленнєвих сигналів, які працюють на все більш низьких швидкостях (тобто забезпечують більшу ступінь стиснення мовленнєвої інформації). Основними причинами, за якими виникає потреба у стисненні мовленнєвих сигналів, є обмежена пропускна здатність типових каналів зв'язку і необхідність криптографічного захисту інформації. Розроблення методів ефективного кодування мовленнєвих сигналів вимагає застосування простих і в той же час ефективних параметричних моделей мовленнєвих сигналів, які базуються на фізичних і фізіологічних процесах їхнього утворення, а також особливостях слухового сприйняття. До відомих моделей належать модель голосового тракту через поєднання акустичних труб, а також моделі із змішаним збудженням. Проте сьогодні найбільш поширеною є авторегресивна (АР) модель утворення мовленнєвих сигналів [1], в якій вони моделюються як результат проходження керувального (збуджувального) процесу через АР фільтр

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

де $s(n)$ — мовленнєвий сигнал; $w(n)$ — збуджувальний процес; g — коефіцієнт

підсилення; $a_k, k = 1,2,\dots, p$, — АР коефіцієнти. Порядок АР моделі p вибирається зазвичай рівним 10.

Основною властивістю параметрів АР моделі є те, що вони відносно повільно змінюються з часом і залишаються майже незмінними на відрізках (фреймах) тривалістю приблизно 20 мс. Таким чином, завдання кодування поділяють на дві основні підзадачі, що виконуються на кожному фреймі мовленнєвого сигналу: кодування збуджувального процесу (джерела) і кодування АР параметрів a_k, $k = 1,2,\dots, p$.

Відомо, що за відсутності фонових шумів параметри АР моделі (1) обчислюють за допомогою методів лінійного прогнозування [1], виходячи із мінімізації критерію

$$\min_{a_{1:p}} \sum_{l=1}^{L} (s(l) + \sum_{k=1}^{p} a_k s(l-k))^2.$$

Таким чином маємо систему із p лінійних рівнянь [1]. Проте, як вже було зазначено, ситуація ускладнюється за наявності шуму v, коли відомі лише зашумлені спостереження z:

$$z(n) = s(n) + v(n). \tag{2}$$

Щоб знайти оцінки АР коефіцієнтів зазвичай використовують принцип максимальної правдоподібності, що зумовлює необхідність пошуку максимумів відповідних багатоекстремальних функціоналів [2]. Функція правдоподібності сигналу (2)

$$f(Z \mid a_1, \dots, a_p, g) = \frac{1}{\sqrt{\det 2\pi C}} \exp\left(-\frac{1}{2} Z^T C^{-1} Z\right) \tag{3}$$

є умовною гаусівською щільністю розподілу вектора спостережень $Z = [z(1), z(2), \dots, z(L)]$, де C — коваріаційна матриця вектора Z. Принцип максимальної правдоподібності полягає у тому, що за наявності вектора спостережень Z оптимальний набір параметрів максимізує функцію правдоподібності (3).

Більшість наявних завадостійких методів оцінювання АР параметрів є модифікаціями методу Expectation-Maximization (EM), що дає змогу отримати оцінку максимальної правдоподібності [3]. Точна реалізація цього методу оцінювання АР параметрів була наведена в [2] і може вважатися еталоном для порівняння з іншими методами. Загальною проблемою ітеративного підходу методу EM є необхідність якісного початкового наближення. Для низьких відношень сигнал–шум (ВСШ) неякісна ініціалізація методу EM може призвести до локального мінімуму замість локального максимуму [2].

Зауважимо також, що метод [2] характеризувався високими обчислювальними витратами, оскільки всі розрахунки виконувалися в часовій області.

З огляду на викладене, у цій роботі пропонується видозмінити методику оцінювання АР параметрів заміною моделі АР сигналу в часовій області моделлю в частотній області, що істотно скоротить обчислювальні витрати під час реалізації методу EM. До того ж запропонований метод має таку саму точність обчислень, що і метод у часовій області.

МОДЕЛЬ АВТОРЕГРЕСИВНОГО СИГНАЛУ В ЧАСТОТНІЙ ОБЛАСТІ

Спочатку наведемо векторну форму зашумленого АР сигналу (1) у часовій області

$$Z(k) = S(k) + V(k) = gA^{-1}W(k) + V(k), \tag{4}$$

де $S(k)$, $W(k), V(k), Z(k)$ — блоки із L значень відповідно мовленнєвого сигналу, збуджувального процесу, шуму і спостережень на k-му фреймі:

$$S(k) = [s((k-1)L+1)\ s((k-1)L+2)\ \dots\ s(kL)]^{\mathrm{T}},$$

$$W(k) = [w((k-1)L+1)\ w((k-1)L+2)\ \dots\ w(kL)]^{\mathrm{T}},$$

$$V(k) = [v((k-1)L+1)\ v((k-1)L+2)\ \dots\ v(kL)]^{\mathrm{T}},$$

$$Z(k) = [z((k-1)L+1)\ z((k-1)L+2)\ \dots\ z(kL)]^{\mathrm{T}}.$$

Відбілююча тепліцева матриця заповнена вихідними АР коефіцієнтами моделі (1)

$$A_{ij} = \begin{cases} 1, & i = j, \\ a_{i-j}, & j < i \le j + p, \\ 0, & i < j \,|\, i > j + p. \end{cases}$$

Вважається, що фоновий шум є кольоровим (colored noise) і характеризується власними АР коефіцієнтами $b_1, \dots, b_q$, тобто вектор V можна представити як $V(k) = hB^{-1}U(k)$, де підсилення h і відбілююча матриця B мають той самий сенс, що g і A для мовленнєвого сигналу.

У цій роботі пропонується замінити (4) моделлю у частотній області

$$Z_\omega(n) = S_\omega(n) + V_\omega(n) = \frac{gW_\omega(n)}{A_\omega(n)} + \frac{hU_\omega(n)}{B_\omega(n)}, \tag{5}$$

де нижні індекси ω позначають перетворення Фур'є вихідних масивів:

$$Z_\omega(n) = \sum_{l=1}^{L} z(l) e^{-2\pi i l n/L}, n = 1, \dots, L;$$

$$S_\omega(n) = \sum_{l=1}^{L} s(l) e^{-2\pi i l n/L}, n = 1, \dots, L;$$

$$V_\omega(n) = \sum_{l=1}^{L} v(l) e^{-2\pi i l n/L}, n = 1, \dots, L;$$

$$W_\omega(n) = \sum_{l=1}^{L} w(l) e^{-2\pi i l n/L}, n = 1, \dots, L;$$

$$U_\omega(n) = \sum_{l=1}^{L} u(l) e^{-2\pi i l n/L}, n = 1, \dots, L.$$

Аналогічно масиви $A_\omega(n)$, $B_\omega(n)$ визначають у такий спосіб:

$$A_\omega(n) = 1 + \sum_{k=1}^{p} a_k e^{-2\pi i k n/L}, n = 1, \dots, L;$$

$$B_\omega(n) = 1 + \sum_{k=1}^{q} b_k e^{-2\pi i k n/L}, n = 1, \dots, L.$$

Дійсно, модель у частотній області (5) еквівалентна заміні в моделі (4) тепліцевих матриць A і B на відповідні їм циркулятні матриці [4].

На основі моделі (5) функцію правдоподібності АР параметрів записують у вигляді

$$p(Z_\omega / \theta) = (2\pi)^{-0.5L} [\det \mathrm{Cov}\{Z_\omega, Z_\omega\}]^{-0.5} \exp[-0.5 Z_\omega^T \mathrm{Cov}^{-1}\{Z_\omega, Z_\omega\} Z_\omega], \tag{6}$$

де вектор θ містить АР параметри сигнала і шума:

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

На основі формули (5) отримуємо вираз для коваріаційної матриці спостережень у частотній області

$$\mathrm{Cov}\{Z_\omega, Z_\omega\} = L\,\mathrm{diag}\{P_s + P_v\}, \tag{7}$$

де P_s — квадрат спектра (амплітудно-частотної характеристики) АР фільтра з коефіцієнтами $a_k, k = 1, \ldots, p$, і підсиленням g:

$$P_s(n) = \frac{g^2}{|A_\omega(n)|^2}, n = 1, \ldots, L. \tag{8}$$

Аналогічно визначаємо масив P_v:

$$P_v(n) = \frac{h^2}{|B_\omega(n)|^2}, n = 1, \ldots, L. \tag{9}$$

На відміну від представлення в часовій області (4), коваріаційна матриця (7) є діагональною, що спрощує подальше обчислення.

Підставляючи (7)–(9) у (6), отримуємо

$$p(Z_\omega / \theta) = (2\pi)^{-0.5L}\left[\prod_{n=1}^{L}\delta_n\right]^{-0.5}\exp\left[-0.5\sum_{n=1}^{L}\delta_n|Z_\omega(n)|^2\right], \tag{10}$$

де вагові коефіцієнти $\delta_n, n = 1, \ldots, L$, визначають у такий спосіб:

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

Зручнішим для запису є логарифм функції правдоподібності (ФП):

$$\log p(Z_\omega / \theta) = \gamma - 0.5\sum_{n=1}^{L}\log\delta_n - 0.5\sum_{n=1}^{L}\delta_n|Z_\omega(n)|^2, \tag{11}$$

де константа γ не впливає на його максимізацію.

Таким чином, формули (10) та (11) дають змогу економічно обчислити функцію правдоподібності та її логарифм. Це може бути застосовано, зокрема, для максимізації функції правдоподібності за допомогою методів, що не використовують обчислення похідних цільової функції або для пошуку початкового наближення для градієнтних методів.

МЕТОД EXPECTATION-MAXIMIZATION У ЧАСТОТНІЙ ОБЛАСТІ

Як вже зазначалось, алгоритм ЕМ дає змогу знайти локальний максимум функції правдоподібності за умови якісного вибору початкового наближення. Формалізм методу ЕМ вимагає визначення так званих «неповних» і «повних» даних [2, 3]. Покладемо Z_ω як неповні і вектор X як повні дані: $X = [S_\omega; V_\omega]$.

На E-кроці методу ЕМ обчислюється апостеріорне значення логарифма ФП вектора повних даних:

$$Q(\theta, \hat{\theta}^{(l)}) = E\{\log p(X / \theta) / Z_\omega, \hat{\theta}^{(l-1)}\}, \tag{12}$$

де l — номер ітерації методу ЕМ.

Вираз для $\log p(X / \theta)$ можна отримати аналогічно формулі (11). Використовуючи співвідношення [1]

$$\sum_{n=1}^{L}\log|A_\omega(n)| \approx 0, \sum_{n=1}^{L}\log|B_\omega(n)| \approx 0$$

та опускаючи константи, які не залежать від θ, отримуємо:

$$\log p(X / \theta) = -L \log g - 0.5 g^{-2} \sum_{n=1}^{L} |S_\omega(n)|^2 |A_\omega(n)|^2 - L \log h - \\ - 0.5 h^{-2} \sum_{n=1}^{L} |V_\omega(n)|^2 |B_\omega(n)|^2 . \quad (13)$$

Підставляючи (13) у (12), отримуємо

$$E\{\log p(X / \theta) / Z_\omega\} = -L \log g - 0.5 g^{-2} \sum_{n=1}^{L} E\{|S_\omega(n)|^2 / Z_\omega\} |A_\omega(n)|^2 - L \log h - \\ - 0.5 h^{-2} \sum_{n=1}^{L} E\{|V_\omega(n)|^2 / Z_\omega\} |B_\omega(n)|^2 . \quad (14)$$

Вирази для $E\{|S_\omega(n)|^2 / Z_\omega\}$, $E\{|V_\omega(n)|^2 / Z_\omega\}$ можна отримати за формулою для апостеріорної кореляції [5]:

$$E\{|S_\omega(n)|^2 / Z_\omega\} = LP_sP_v / (P_s + P_v) + P_s^2 |Z_\omega|^2 / (P_s + P_v)^2, \quad (15)$$

$$E\{|V_\omega(n)|^2 / Z_\omega\} = LP_sP_v / (P_s + P_v) + P_v^2 |Z_\omega|^2 / (P_s + P_v)^2 . \quad (16)$$

У цьому разі спектри P_s, P_v визначають згідно з (8), (9) через оцінки параметрів із попередньої ітерації.

На M-кроці методу ЕМ максимізується вираз (14). Диференціювання (14) за $a_1, \ldots, a_p, g$ приводить до співвідношення щодо оцінок АР параметрів:

$$\sum_{j=1}^{p} r_s(i-j) a_j = -r_s(i), i = 1, \ldots, p, \quad (17)$$

$$g = \sqrt{\frac{r_s(0) + \sum_{k=1}^{p} a_k r_s(k)}{L}},$$

де коефіцієнти $r_s(k), k = 0, \ldots, p$, є першими $p+1$ елементами оберненого перетворення Фур'є масиву (15).

Аналогічно формули для оцінок АР параметрів шума мають вигляд

$$\sum_{j=1}^{q} r_v(i-j) b_j = -r_v(i), i = 1, \ldots, q, \quad (18)$$

$$h = \sqrt{\frac{r_v(0) + \sum_{k=1}^{q} b_k r_v(k)}{L}},$$

де коефіцієнти $r_v(k), k = 0, \ldots, q$, є першими $q+1$ елементами оберненого перетворення Фур'є масиву (16).

Системи (17) і (18) є тепліцевими і можуть бути розв'язані рекурсивним методом Левінсона–Дарбіна [1]. Деталі реалізації цієї процедури з урахуванням специфіки процесорів із фіксованою комою наведені в [6].

Таким чином, повторюючи ітеративно E- та M-кроки, отримуємо шукані оцінки АР параметрів мовленнєвого сигналу $a_1, \ldots, a_p, g$ і шуму $b_1, \ldots, b_q, h$.

Обчислювальні витрати оптимізованого методу ЕМ. Оскільки основними операціями методу є швидкі перетворення Фур'є (ШПФ), загальна кількість операцій істотно залежить від кількості машинних циклів, необхідних для одного ШПФ. Надалі припускаємо, що кількість операцій для оброблення фрейму довжиною L становить $N_{FFT} = \alpha L \log_2 L$ (якщо L є ступенем двох), де α —

коефіцієнт, що залежить від типу обчислювальної платформи. Розглядаються три можливих значення α: 3 (теоретичне значення [4]), 1 і 0.3 (досяжне завдяки апаратній реалізації ШПФ на цифрових сигнальних процесорах). Загальні оцінки обчислювальних витрат та методика обчислення ШПФ наведені в [7, 8].

Загальна кількість операцій в секунду для запропонованого методу оцінювання в частотній області становить

$$N = [\alpha \log_2 L + 2p + 12] K f_s. \qquad (19)$$

Це означає, що для базової конфігурації: $K = 5$ ітерацій ЕМ, частота дискретизації 8000 Гц, порядки АР моделей $p = 10$, $q = 8$, довжина фрейму $L = 256$ (32 мс) і $\alpha = 1$, величина (19) становить приблизно 1.6 MFLOPS (мільйонів операцій в секунду).

У той же час витрати методу, що працює в часовій області, становлять

$$N' = [4(p^2 + q^2) + 2p^3 / 3] K f_s.$$

Відношення N' / N становить 38.4 для $\alpha = 0.3$, 33.1 для $\alpha = 1$ і 23.6 для $\alpha = 3$. Отже, підхід у частотній області дає змогу скоротити число операцій від 23.6 до 38.4 раза для різних α.

ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Необхідно було з'ясувати, чи можна за допомогою методів, що ґрунтуються на часовій (4) і частотній (5) моделях АР сигналу, отримати еквівалентні оцінки АР параметрів. В експериментах використовувалися збалансовані речення, які промовляли чотири диктори-чоловіки і двоє дикторів-жінок і були оцифровані з частотою 8000 Гц. Ці сигнали були змішані з адитивним кольоровим шумом за різних ВСШ: 5, 10, 15 і 20 дБ. Оцінювання АР параметрів здійснювалося на фреймах тривалістю 32 мс (L = 256 відліків) для порядків АР моделей сигналу і шуму $p = 10$ і $q = 8$ відповідно.

Були проаналізовані метод ЕМ в часовій області та запропонована в цій роботі модифікація методу ЕМ, що ґрунтується на моделі в частотній області.

Для реалізації методу ЕМ використовувалися $K = 5$ його ітерацій. Методи порівнювалися за такими критеріями:

— середньоквадратичне відхилення спектрів;

— міра Ітакури–Саіто [1];

— нормована міра Ітакури–Саіто [1].

З наведених у табл. 1 оцінок АР параметрів випливає, що запропонована модифікація методу ЕМ забезпечує майже однакову з методом ЕМ у часовій області точність усіх проаналізованих критеріїв.

Зауважимо, що обчислення АР коефіцієнтів із залученням фільтра Вінера в частотній області використовувались у [9], проте згаданий метод мав наближений характер і не забезпечував достатню якість оцінювання параметрів у порівнянні із методом у часовій області, що був наведений у [2].

Таблиця 1

ВСШ, дБ	Міра спектрального спотворення		Міра Ітакури–Саіто		Нормована міра Ітакури–Саіто	
	Часова область	Частотна область	Часова область	Частотна область	Часова область	Частотна область
5	3.77	3.79	0.79	0.78	0.36	0.37
10	3.10	3.10	0.48	0.48	0.26	0.26
15	2.50	2.47	0.33	0.30	0.18	0.18
20	2.00	1.90	0.24	0.20	0.12	0.11

ВИСНОВКИ

1. У статті розглянуто задачу оцінювання АР параметрів мовленнєвих сигналів за умови наявності адитивного шуму.

2. На основі представлення АР сигналу в частотній області наведено спосіб обчислення функціоналу правдоподібності АР параметрів. Цей спосіб, зокрема, може застосовуватися для максимізації функції правдоподібності за допомогою методів, що не використовують обчислення похідних цільової функції.

3. Наведено реалізацію алгоритму Expectation-Maximization для оцінювання АР параметрів у частотній області.

4. Аналіз різних мір спотворення мовленнєвих сигналів показав, що запропоновані підходи в частотній області мають таку саму точність, що і відповідні їм підходи в часовій області, але характеризуються суттєво меншими обчислювальними витратами (від 23.6 до 38.4 раза для різних модифікацій обчислення швидкого перетворення Фур'є).

СПИСОК ЛІТЕРАТУРИ

1. Маркел Дж.Д., Грей А.Х. Линейное предсказание речи. Москва: Радио и связь, 1980. 307 с.
2. Dempster A., Lair N., Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statistic. Soc.* 1977. Vol. 39. P. 1–38.
3. Gannot S., Burnstein D., Weinstein E. Iterative and sequential Kalman filter-based speech enhancement algorithms. *Transactions Speech Audio Processing*. 1998. Vol. 6. P. 373–385.
4. Воеводин В.В., Тыртышников Е.Е. Вычислительные процессы с теплицевыми матрицами. Москва: Наука, 1987. 320 с.
5. Сейдж Э., Мелс Дж. Теория оценивания и ее применение в связи и управлении. Серия: Статистическая теория связи. Москва: Связь, 1976. 496 с.
6. Семенов В.Ю. Методы вычисления и кодирования параметров авторегрессионной модели речи при разработке вокодера на базе сигнального процессора с фиксированной точкой. *Проблемы управления и информатики.* 2019. № 1. С. 41–50.
7. Задирака В.К., Мельникова С.С. Цифровая обработка сигналов. Київ: Наук. думка, 1993. 294 с.
8. Задирака В.К. Теория вычисления преобразования Фурье. Київ: Наук. думка, 1983. 216 с.
9. Lim J.S., Oppenheim A.V. All–pole modeling of degraded speech. *IEEE Trans. Acoust. Speech Signal Proces.* 1978. Vol. 26. P. 197–210.

V.K. Zadiraka, V.Yu. Semenov, Ye.V. Semenova

METHOD OF NOISE-ROBUST ESTIMATION OF PARAMETERS OF AUTOREGRESSIVE MODEL IN FREQUENCY DOMAIN

Abstract. The article considers the problem of estimating the parameters of the autoregressive (AR) signal in the presence of background noise. Based on the frequency representation of the AR signal, a technique of calculating the likelihood function of the AR parameters is shown and the implementation of the expectation-maximization method for iterative evaluation of the AR parameters is considered. Analysis of different measures of distortion of speech signals shows that the proposed approaches in the frequency domain have the same accuracy with the corresponding approaches in the time domain, but are characterized by significantly lower computing costs.

Keywords: autoregressive model, likelihood function, Expectation-Maximization method, fast Fourier transform.

Надійшла до редакції 13.04.2021