
ТЕОРИЯ И МЕТОДЫ ОБРАБОТКИ СИГНАЛОВ

УДК 621.391:004.934

МЕТОД АВТОРЕГРЕССИОННОГО МОДЕЛИРОВАНИЯ РЕЧЕВОГО СИГНАЛА С ИСПОЛЬЗОВАНИЕМ ОГИБАЮЩЕЙ ПЕРИОДОГРАММЫ ШУСТЕРА В КАЧЕСТВЕ ОПОРНОГО СПЕКТРАЛЬНОГО ОБРАЗЦА

© 2023 г. В. В. Савченко*

*Национальный исследовательский университет “Высшая школа экономики”,
ул. Б. Печерская, 25, Нижний Новгород, 603155 Российская Федерация*

**E-mail: vvsavchenko@yandex.ru*

Поступила в редакцию 30.06.2022 г.

После доработки 23.09.2022 г.

Принята к публикации 27.09.2022 г.

Рассмотрена задача авторегрессионного моделирования речевого сигнала по данным дискретного преобразования Фурье в режиме скользящего окна наблюдений небольшой длительности (миллисекунды). Исследована проблема устойчивости формируемой авторегрессионной модели. Предложено для ее преодоления использовать в качестве опорного спектрального образца огибающую периодограммы Шустера. Разработан новый метод авторегрессионного моделирования, в котором детектирование спектральной огибающей осуществляется с использованием рециркулятора последовательности отсчетов в частотной области. Рассмотрен пример его практической реализации, поставлен и проведен натурный эксперимент. По результатам эксперимента сделаны выводы о достижении существенно-го выигрыша в отношении не только устойчивости, но и точности авторегрессионной модели речевого сигнала.

DOI: 10.31857/S0033849423020122, EDN: LDATIU

ВВЕДЕНИЕ

На протяжении нескольких десятков лет авторегрессионная (АР) модель привлекает к себе внимание исследователей и практиков в качестве способа эффективного кодирования речевого сигнала с многократным сжатием данных в информационных системах и технологиях самого разного назначения: от передачи речи по цифровым каналам связи до речевой аналитики и голосового управления [1, 2]. Как следствие, построение АР-модели, или АР-моделирование давно стало классической задачей в области автоматической обработки речи (АОР) [3–5]. Неудивительно поэтому, что существует множество разных подходов к ее решению, в рамках которых предложено множество разных методов и алгоритмов. Среди них доминируют параметрические методы цифрового спектрального анализа [6]. Примером может служить метод Берга, в котором гарантируется устойчивость формируемой АР-модели речевого сигнала вне зависимости от его длительности и тонкой структуры [7]. Однако точность метода Берга далеко не всегда отвечает требованиям практики, в том числе из-за эффектов нелинейной обработки: смещения и расщепления спек-

тральных мод в условиях малых выборок наблюдений [8].

Для преодоления недостатков параметрических методов был предложен [9] метод спектрального АР-моделирования речевого сигнала в режиме скользящего окна наблюдений конечной длительности τ . Благодаря использованию в этом методе дискретного преобразования Фурье (ДПФ), основанного на линейной обработке речевого сигнала, его авторам удалось существенно повысить точность АР-модели. Определенные сомнения при этом вызывает проблема устойчивости или стабильности [10] формируемой модели в смысле цифровой рекурсивной фильтрации [11]. Причем ее острота резко возрастает по мере увеличения требований пользователя к результатам итеративной вычислительной процедуры [9]. Решению указанной проблемы и посвящена, главным образом, данная статья.

Статья написана по результатам проведенного автором исследования в развитие результатов его предыдущих работ, опубликованных в соавторстве с сотрудниками Лаборатории алгоритмов и технологий анализа сетевых структур НИУ “Высшая школа экономики”, Нижний Новгород, Россия за период

2020–2022 гг. Цель исследования – разработка нового метода АР-моделирования речевого сигнала повышенной точности и устойчивости в условиях малых выборок.

1. ПОСТАНОВКА ЗАДАЧИ

Задача формулируется как оптимизационная в терминах дискретного спектрального моделирования (ДСМ) [12]. Пусть задан речевой фрейм $\{x(n)\}$ конечной размерности $N = \tau F$, где F – частота дискретизации речевого сигнала. Отталкиваясь от общего выражения на выделенном наборе из $M \gg 1$ дискретных частот $f_m = f_{m-1} + \Delta f$, $\Delta f = F/M$, спектральной плотности мощности (СПМ) АР-модели p -го порядка [8]

$$G_p(f_m; \mathbf{b}_l) = \frac{\sigma_0^2 T}{\left| \sum_{i=0}^p b_l(i) \exp(-j2\pi i f_m T) \right|^2}, \quad m = 0, M-1 \quad (1)$$

($T = 1/F$, $\sigma_0^2 = \text{const}$), путем последовательных приближений (итераций)

$$\forall i \leq p: b_l(i) = b_{l-1}(i) - \gamma_0 \frac{d}{db_{l-1}(i)} \rho(\mathbf{b}_{l-1}), \quad l = 1, 2, \dots, \quad (2)$$

с шагом $\gamma_0 = \text{const}$ найдем оптимальный вектор весовых коэффициентов $\mathbf{b}_l = \{b_l(i)\}$ из условия минимизации модифицированной величины COSH-расстояния:

$$\rho(\mathbf{b}_l) = \sqrt{\left[M^{-1} \sum_{m=0}^{M-1} G_0(f_m) G_p^{-1}(f_m; \mathbf{b}_l) \right] \left[M^{-1} \sum_{m=0}^{M-1} G_p(f_m; \mathbf{b}_l) G_0^{-1}(f_m) \right]} - 1 \geq 0, \quad (3)$$

формируемой СПМ (1) относительно ее опорного образца $\{G_0(f_m)\}$ в теоретико-информационном смысле [9, 13]. Ввиду широкого распространения АР-модели в области АОР [7, 8] задача в данной постановке представляет очевидный интерес. Кроме того, из справедливости тождества

$$\rho(\mathbf{b}_l)_{|G_p(f_m; \mathbf{b}_l)=G^*(f_m)} = \rho(\mathbf{b}_l)_{|G_p(f_m; \mathbf{b}_l)=cG^*(f_m)}$$

для произвольной СПМ $G^*(f) > 0$ и неотрицательной константы $c > 0$, в том числе при равенстве $c = 1/\sigma_0^2$, вытекает инвариантность меры (3) к масштабному множителю σ_0^2 . Это исключает необходимость пошагового масштабирования АР-модели (1) в рамках итеративной вычислительной процедуры (2).

В работе [9] в качестве спектрального образца $\{G_0(f_m)\}$ использовалась основанная на M -точечном ($M \geq N$) ДПФ речевого сигнала ($j = \sqrt{-1}$)

$$S_x(jf_m) = T \sum_{n=0}^{N-1} x(n) \exp(-j2\pi n f_m T) = T \sum_{n=0}^{N-1} x(n) \exp(-j2\pi n m M^{-1}) \quad (4)$$

периодограмма Шустера общего вида [8]

$$G_x(f_m) = (MT)^{-1} |S_x(jf_m)|^2, \quad m = 0, 1, \dots, M-1. \quad (5)$$

Однако в этом случае не гарантируется устойчивость формируемой АР-модели (1), и это острая проблема для большинства известных методов ДСМ [10, 12].

Эффективным способом решения указанной проблемы представляется использование вместо

периодограммы (5) ее огибающей $\{\bar{G}_x(f_m)\}$ [14]. Во-первых, этим учитывается известная нацеленность АР-модели на описание именно огибающей СПМ, если мы имеем в виду вокализованный речевой сигнал [7]. И, во-вторых, использование огибающей периодограммы Шустера сильно уменьшает степень изрезанности формы опорного спектрального образца $\{G_0(f_m)\}$ как главного фактора неустойчивости АР-модели в задачах ДСМ [10]. Рассмотрим этот способ подробно в рамках общей формулировки (1)–(3) оптимизационной задачи.

2. СИНТЕЗ АЛГОРИТМА

Согласно акустической теории речеобразования [1, 7] СПМ вокализованных отрезков речевого сигнала имеет квазилинейчатую структуру и состоит из модулированных по амплитуде гармоник частоты основного тона F_0 [15]. Их огибающая содержит важную информацию о внутрипериодной (внутри периода основного тона) тонкой структуре речевого сигнала и именно в этом качестве представляет первостепенный интерес [16]. Аналогичную линейчатую структуру со сдвигом своих отсчетов в частотной области на $\Delta f = F/M$ имеет и периодограмма (5). Причем в силу линейности ДПФ (4) ее огибающая $\{\bar{G}_x(f_m)\}$ по форме подобна огибающей истинной СПМ. Вместе с тем АР-модель (1) является наиболее распространенным способом ее математического описания в задачах АОР [1]. Желательно, разумеется, при этом использовать ее наиболее точный оптимальный вариант.

Однако здесь возникает серьезная проблема в отношении критерия оптимальности. Все сколько-нибудь известные на данный момент критерии – Акаике, Шварца и их разнообразные модификации [17] – рассчитаны исключительно на эргодические случайные процессы. Речевой сигнал, рассматриваемый в условиях малых выборок, к ним не относится. Выход подсказывает работа [15], в которой предложен алгоритм оптимизации порядка $p < \infty$ АР-модели (1) через решение обратной задачи речеобразования по принципу “от речевого сигнала к его голосовому источнику”.

Вокализованная речь имеет своим источником периодическую последовательность коротких (доли миллисекунды) импульсов возбуждения голосового тракта диктора [16]. Обозначим ее СПМ через $\{G_y(f_m)\}$. В предположении о линейности голосового тракта и заданной априори СПМ речевого сигнала $\{G^*(f_m)\}$ в идеале будем иметь [5]

$$G_y(f_m) = cG^*(f_m)/\bar{G}^*(f_m), \quad m = \overline{0, M-1}. \quad (6)$$

Это известное уравнение выравнивающего [18] фильтра в частотной области.

Идеальная СПМ (6) отличается от истинной $\{G^*(f_m)\}$ только формой огибающей $\{\bar{G}_y(f_m)\}$. В силу малой длительности импульсов возбуждения она имеет практически прямоугольный вид в полосе рабочих частот $f_m \leq 0.5F$ [11]. Для всех $m < M$ при этом выполняется равенство

$$\begin{aligned} \bar{G}_y(f_m) &= c\bar{G}^*(f_m)/\bar{G}^*(f_m) = \\ &= c\bar{G}^*(f_m)/\bar{G}^*(f_m) = c = \text{const.} \end{aligned}$$

на роль целевой функции (3) рассматриваемой оптимизационной задачи. В ней в качестве опорного спектрального образца $\{G_0(f_m)\}$ используется огибающая $\{\bar{G}_x(f_m)\}$ периодограммы Шустера. Задача переходит в практическую плоскость: демодуляция СПМ (5) или оценка спектральной огибающей. Один из перспективных вариантов ее решения основывается на применении рециркулятора последовательности отсчетов в частотной области.

3. ПРИМЕР ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ

По определению (5) периодограмма Шустера является неотрицательной функцией частоты. Поэтому ее демодуляция, или амплитудное детектирование в частотной области, не требует обычной нелинейной (диодной) обработки сиг-

Отсюда в соответствии с (3) получим тождество

$$\begin{aligned} &\sqrt{\left[M^{-1} \sum_{m=0}^{M-1} c^{-1} \bar{G}_y(f_m)\right] \left[M^{-1} \sum_{m=0}^{M-1} c \bar{G}_y^{-1}(f_m)\right]} - 1 = \\ &= \sqrt{\left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_y(f_m)\right] \left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_y^{-1}(f_m)\right]} - 1 \equiv 0. \end{aligned} \quad (7)$$

А это признак идеального выравнивающего фильтра (6). Правда, данное тождество не достижимо в строгом смысле в условиях априорной неопределенности в отношении истинного вида СПМ $\{G^*(f_m)\}$. Однако его можно использовать в качестве критерия точности АР-модели (1). Для этого подставим в (6) вместо неизвестных в общем случае СПМ $\{G^*(f_m)\}$ и $\{\bar{G}^*(f_m)\}$ их текущие выборочные оценки: $\{\hat{G}^*(f_m)\}$ и $\{\hat{\bar{G}}^*(f_m)\}$ соответственно. В рамках формализации (1)–(5) в результате получим равенство

$$\begin{aligned} \hat{G}_y(f_m) &= \hat{G}^*(f_m)/\hat{\bar{G}}^*(f_m) = \\ &= G_x(f_m)/G_p(f_m; \mathbf{b}_l) \triangleq G_y(f_m; \mathbf{b}_l), \quad m = \overline{0, M-1}. \end{aligned}$$

Уравнение выравнивающего фильтра в данном случае определяется через АР-модель (1). Отсюда в первом приближении по аналогии с равенством (6) можно записать

$$\bar{G}_y(f_m; \mathbf{b}_l) = \bar{G}_x(f_m)G_p^{-1}(f_m; \mathbf{b}_l), \quad m = \overline{0, M-1}. \quad (8)$$

После подстановки выражения (8) в левую часть тождества (7) получим меру качества АР-модели (1) вида

$$\rho(\mathbf{b}_l) = \sqrt{\left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_x(f_m)G_p^{-1}(f_m; \mathbf{b}_l)\right] \left[M^{-1} \sum_{m=0}^{M-1} G_p(f_m; \mathbf{b}_l)\bar{G}_x^{-1}(f_m)\right]} - 1 \quad (9)$$

нала в целях выпрямления его полярности. Достаточно сгладить последовательность отсчетов речевого сигнала согласно разностному уравнению [11]

$$\begin{aligned} \bar{G}_x(f_m) &= \beta_x \bar{G}_x(f_m - \Delta f) + G_x(f_m), \\ m &= 1, \dots, M-1, \end{aligned} \quad (10)$$

где $\beta_x = \text{const}$. Выражение (10) описывает работу стандартного рециркулятора в функции межпериодного накопителя последовательности отсчетов речевого сигнала в режиме скользящего окна наблюдений в частотной области. В определенном смысле это аналог периодограммы Даньелла со сглаживанием соседних отсчетов [8]. Размер (длина) окна, или инерционность рециркулятора, регулируется выбором значения его параметра $\beta_x > 0$. Показатель инерционности [11]

$$\theta = -\Delta f / \ln \beta_x = -FM^{-1} / \ln \beta_x$$

ограничен снизу величиной $\Delta f_0 = F/N$ разрешающей способности ДПФ (4) по частоте. Однако выражение (8) накладывает на показатель θ естественные ограничения и сверху – в расчете на определенную степень гладкости формы АР-оценки СПМ $\{G_p(f_m; \mathbf{b}_l)\}$. Поэтому ограничимся равенством $\theta = \Delta f_0$, откуда получим верхнюю границу

$$\beta_x = \exp(-\Delta f / \Delta f_0) = \exp(-NM^{-1}). \quad (11)$$

Выражения (10) и (11) в совокупности определяют огибающую периодограммы Шустера при ее использовании в (12) в качестве опорного спектрального образца. Например, при $M = N = 128$ будем иметь $\beta_x \leq 0.37$.

Наглядной иллюстрацией к этому примеру служат рис. 1, на котором представлены три графика: периодограммы (5) и ее огибающей (10) для двух значений параметра β_x . Графики получены в границах одного и того же фрейма гласного звука речи “а” при равенствах $\tau = 16$ мс, $F = 8$ кГц и $M = 128$. Как видим, варьируя значение коэффициента усиления β_x , мы можем в широких пределах регулировать форму спектральной огибающей $\{\bar{G}_x(f_m)\}$ при сохранении достаточно высокой степени ее подобия форме исходной периодограммы (5). Система равенств

$$\forall m < M: G_0(f_m) = \bar{G}_x(f_m) \quad (12)$$

и выражения (9)–(11) в совокупности определяют новый метод АР-моделирования речевого сигнала по данным его ДПФ на интервале наблюдений конечной длительности τ . Эффективность данного метода была исследована экспериментально.

4. ПРОГРАММА И МЕТОДИКА ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА

Объектом исследования служили сигналы шести русских гласных фонем из предыдущей работы автора [9]. Идея состояла в сравнительном анализе эффективности двух методов АР-моделирования речи: предложенного в данной работе – с использованием спектральной огибающей (12), и его прототипа, в котором в качестве опорного спектрального образца использовалась периодограмма Шустера (5). Напомним, что длительность каждого сигнала $T_x = 2...3$ с изначально предполагала его линейное членение на последовательность коротких ($\tau = 16$ мс) фреймов при их частичном (по 3 мс в начале и в конце) взаимном перекрытии во времени. В результате для каждой фонемы была создана представительная речевая база данных объемом $K = 1.6T_x/\tau = 200...300$ однородных фреймов $\{x(n)\}$. При частоте дискретизации речевого сигнала $F = 8$ кГц размерность каждого из них составила $N = 128$.

В дальнейшем по каждому фрейму с использованием 128-точечного ДПФ ($M = 128$) была получена периодограмма (5), а после этого – согласно выражениям (10)–(12) при равенстве $\beta_x = 0.25$ – ее огибающая $\{\bar{G}_x(f_m)\}$ в качестве соответствующего спектрального образца $\{G_0(f_m)\}$. Его АР-модель (1) порядка $p = 10$ (типовое значение для систем АОР [1]) была получена согласно итеративной процедуре (2) при ее инициализации вектором параметров $\mathbf{b}_0 = (1, -\mathbf{a}_p)$, где $\mathbf{a}_p$ – вектор коэффициентов линейной авторегрессии p -го порядка. При этом было использовано выражение для градиента критериальной функции (9) общего вида [9]

$$\forall i = \overline{0, p}: \frac{d}{db_l(i)} \rho(\mathbf{b}_l) = M^{-1} \left[\sqrt{\frac{g_1(\mathbf{b}_l)}{g_2(\mathbf{b}_l)}} \sum_{k=0}^p b_l(k) \sum_{m=0}^{M-1} \bar{G}_x(f_m) \cos[2\pi(k-i)f_m T] - \sqrt{\frac{g_2(\mathbf{b}_l)}{g_1(\mathbf{b}_l)}} \sum_{k=0}^p b_l(k) \sum_{m=0}^{M-1} \bar{G}_x^{-1}(f_m) G_p^2(f_m; \mathbf{b}_l) \cos[2\pi(k-i)f_m T] \right], \quad l = 1, 2, \dots,$$

где введены обозначения

$$g_1(\mathbf{b}_l) \triangleq M^{-1} \sum_{m=0}^{M-1} G_p(f_m; \mathbf{b}_l) \bar{G}_x^{-1}(f_m);$$

$$g_2(\mathbf{b}_l) \triangleq M^{-1} \sum_{m=0}^{M-1} \bar{G}_x(f_m) G_p^{-1}(f_m; \mathbf{b}_l).$$

Сходимость итераций (2) в точку $\mathbf{b}^*$ минимума зависимости (9) гарантировалась выбором допустимого (ограниченного сверху) значения шага $\gamma_0 = 0.1$. Однако возникают вопросы принципиального характера: насколько гарантирована устойчивость итоговой АР-модели и какова ее точность?

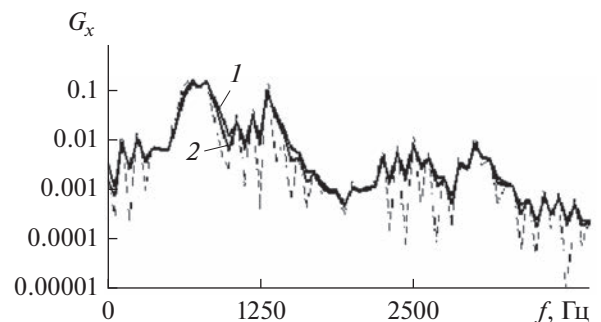


Рис. 1. Графики периодограммы Шустера (штриховая кривая) и ее огибающей (сплошные кривые) в двух вариантах значений параметра β_x : 0.37 (1) и 0.25 (2).

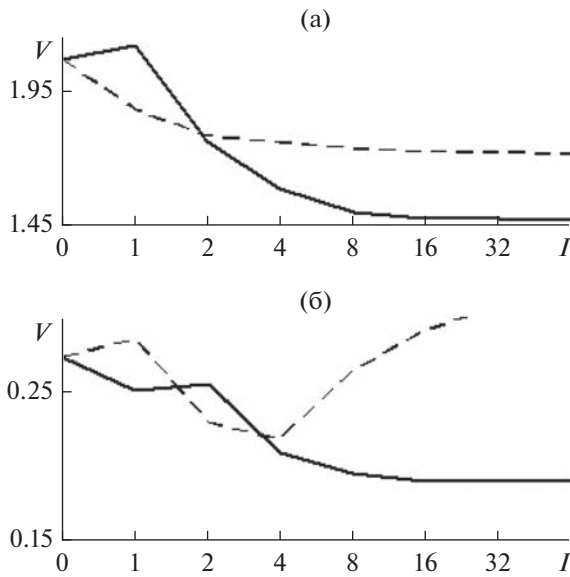


Рис. 2. Кривые сходимости в метрике (3) двух методов АР-моделирования сигнала звука речи “а” (сплошные кривые): с использованием в качестве опорного спектрального образца периодограммы Шустера (а) и ее огибающей (б) в сопоставлении с кривыми сходимости их альтернатив (штриховые кривые): предложенного в статье метода и его прототипа соответственно.

Устойчивость формируемой модели (1) на нулевом шаге итераций обеспечивалась инициализацией вычислительной процедуры (2) с использованием выборочной оценки вектора коэффициентов $\mathbf{a}_p$ методом Берга. Устойчивость ее последующих приближений контролировалась наблюдателем

$$\mathbf{b}_0 = (1; -0.917124236; 0.430847898; -0.265832062; 0.701148961; -0.372491825; -0.002903572; -0.031471774; 0.189665721; 0.140846521; -0.356422549).$$

Как видим, обе кривые быстро сходятся в область оптимума $l > l_0 = 4$, где выполняется приближенное равенство $\mathbf{b}^* \approx \mathbf{b}_8$. В нормированном виде $V_0(l) = V(l)/V(0)$, $l = 1, 2, \dots$, они практически не отличаются друг от друга по динамике (рис. 3). Правда, полученные результаты довольно сильно отличаются между собой по точности. Это хорошо видно из сравнения графиков двух СПМ на

$$\mathbf{b}_8 = (1.153056; -0.939943644; 0.395510652; -0.10847565; 0.719702155; -0.32657449; 0.237992271; 0.034458309; 0.082703038; 0.208791598; -0.258948291).$$

Отметим, что в рамках нового метода был получен качественно иной результат:

$$\mathbf{b}_8 = (1.06302228; -0.868566584; 0.455000405; -0.220341619; 0.718875293; -0.338769231; 0.084934534; -0.007201326; 0.21310955; 0.118502559; -0.305005488).$$

визуально: по виду импульсной характеристики соответствующего формирующего фильтра с комплексным коэффициентом передачи

$$K_p(jf; \mathbf{b}_l) = \frac{b_l(0)}{\sum_{i=0}^p b_l(i) \exp(-j2\pi i f T)} \sim \sqrt{G_p(f; \mathbf{b}_l)}, \quad |f| \leq 0.5F. \quad (13)$$

Рабочей характеристикой, или показателем эффективности предложенного метода служили значения функции (3), рассчитанные согласно выражению (9) для каждого отдельного фрейма данных $\{x(n)\}$. Впоследствии эти значения статистически усреднялись на множестве из K их практически независимых реализаций. При этом погрешность усредненных оценок в ее относительном выражении не вышла за пределы $\delta = 165/\sqrt{K} = 10 \dots 11\%$ при доверительной вероятности, равной 0.9 [15, 19].

5. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Полученные результаты представлены на рисунках. На рис. 2а, 2б показаны кривые сходимости $V(l) = \rho(\mathbf{b}_l)$, $l = 1, 2, \dots$, двух разных методов АР-моделирования речи. Напомним, что различия между рассматриваемыми методами состоят в используемом опорном образце СПМ $\{G_0(f_m)\}$. Первоначально это была периодограмма Шустера (4), (5), а в новом методе — ее огибающая (10). Все остальные их элементы, включая итеративную градиентную процедуру (2), совпадают. Совпадают в данном случае и их начальные приближения вектора АР-параметров по методу Берга:

рис. 4. Кроме того, путем слежения за импульсной характеристикой формирующего фильтра (13) было установлено, что приближения вектора АР-параметров $\mathbf{b}_l$, полученные новым методом (1)–(12), отвечают устойчивому варианту АР-модели (1) для всех $l \leq 64$. Напротив, при применении метода-прототипа устойчивость АР-модели нарушается, начиная, по крайней мере, с восьмого шага приближений, для которого выполняется равенство

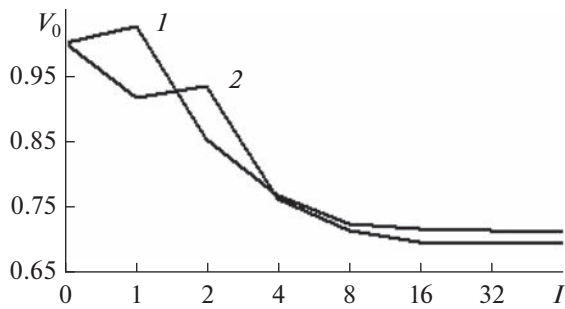


Рис. 3. Нормированные кривые сходимости двух методов АР-моделирования: с использованием в качестве опорного спектрального образца периодограммы Шустера (кривая 1) и ее огибающей (кривая 2).

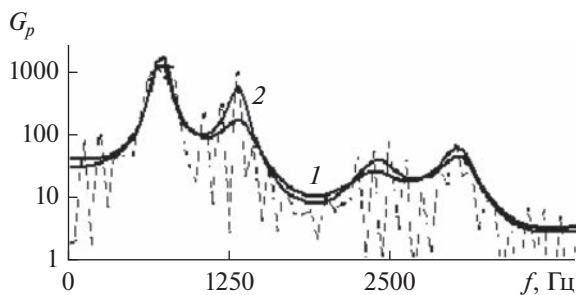


Рис. 4. Семейство СПМ по результатам применения двух методов АР-моделирования: с использованием в качестве опорного спектрального образца периодограммы Шустера (кривая 1) и ее огибающей (кривая 2) в сопоставлении с графиком периодограммы Шустера (штриховая линия).

Наглядной иллюстрацией к сказанному служат штриховые кривые на рис. 2, которые отображают в метрике (3) динамику итераций (2) для альтернативного в каждом из двух случаев метода АР-моделирования. Так, начиная с точки $l > 4$ штриховая кривая на рис. 2б утрачивает все признаки сходимости к оптимуму. Это справедливо для метода-прототипа. Напротив, штриховая кривая на рис. 2а этими признаками определенно обладает. Это относится к новому методу. По-видимому, именно так проявляется себя проблема устойчивости АР-модели (1).

Отметим, что сделанные выводы распространяются и на другие гласные звуки речи контрольного диктора. Подтверждением служат две гистограммы распределений по гласным фонемам средней доли $W(\mathbf{b})$ устойчивых результатов АР-моделирования на рис. 5, где серым фоном отмечены результаты работы [9]. При доверительной вероятности 0.95 погрешность измерений составила в данном случае величину порядка $196/\sqrt{2000} \approx 5\%$ [19]. Таким обра-

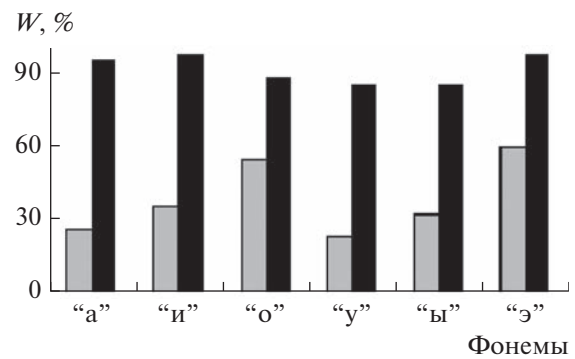


Рис. 5. Гистограммы распределения по гласным фонемам контрольного диктора средней доли устойчивых АР-моделей (1) по результатам четырех итераций (2) при применении двух методов моделирования: с использованием в качестве опорного спектрального образца периодограммы Шустера (серые столбики [9]) и ее огибающей (черные столбики).

зом, если в качестве опорного спектрального образца $\{G_0(f_m)\}$ использовать не непосредственно периодограмму Шустера, а ее огибающую, то практически полностью исключается проблема устойчивости АР-модели в задачах ДСМ.

6. ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Проведенное исследование было бы не полным без ответа на вопрос о сравнительной точности двух представленных на рис. 4 оценках СПМ. По существу, речь идет о критерии качества формируемой АР-модели. Ответом на этот вопрос могут служить результаты второй, заключительной стадии экспериментальных исследований. На этой стадии АР-модель (1) рассматривалась в качестве инверсии квадрата комплексного коэффициента передачи

$$K_p^{-1}(jf; \mathbf{b}_l) \sim \sqrt{G_p^{-1}(f; \mathbf{b}_l)}, |f| \leq 0.5F$$

выравнивающего фильтра (6). СПМ речевого сигнала на выходе этого фильтра

$$G_y(f_m; \mathbf{b}_l) = G_x(f_m) K_p^{-2}(jf_m; \mathbf{b}_l) = c G_x(f_m) G_p^{-1}(f_m; \mathbf{b}_l), m=0, M-1, \quad (14)$$

в идеале должна иметь прямоугольную огибающую $\{\bar{G}_y(f_m; \mathbf{b}_l)\}$, для которой в полосе рабочих частот выполняется тождество (7). В силу неидеальности АР-модели (1) оно приводит к следующему выражению для объективного показателя ее точности:

$$Q(\mathbf{b}_l) = \sqrt{\left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_y(f_m; \mathbf{b}_l) \right] \left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_y^{-1}(f_m; \mathbf{b}_l) \right]} - 1 \geq 1. \quad (15)$$

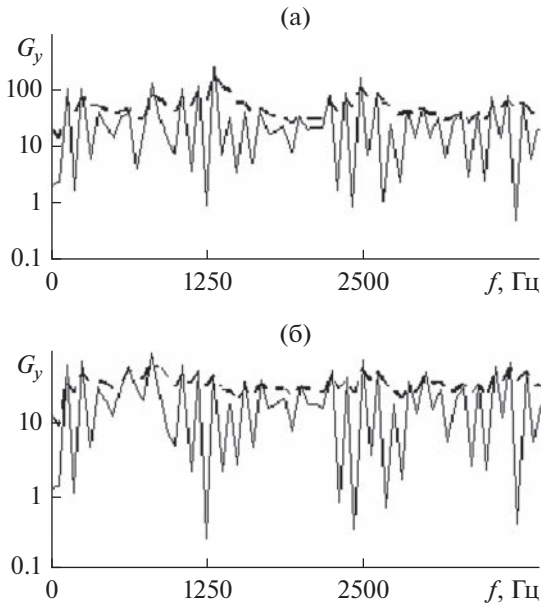


Рис. 6. Графики СПМ сигнала гласного звука речи “а” на выходе выравнивающего фильтра (14) при применении двух методов АР-моделирования (сплошные кривые): метода-прототипа (а) и предложенного нового метода (б), в сопоставлении с графиками двух спектральных огибающих $\bar{G}_y(f_m)$ (штриховые кривые), рассчитанными согласно (16).

Он и служил в дальнейшем характеристикой сравнительной эффективности двух методов АР-моделирования. При этом для выделения спектральной огибающей $\{\bar{G}_y(f_m; \mathbf{b}_l)\}$ по аналогии с (10) использовали рециркулятор

$$\bar{G}_y(f_m) = \beta_y \bar{G}_y(f_m - \Delta f) + G_y(f_m), \quad (16)$$

$$m = 1, \dots, M - 1.$$

Его коэффициент усиления в цепи обратной связи

$$\beta_y = \exp(-\Delta f / F_0) = \exp(-F / (F_0 M)) \approx 0.62$$

рассчитывается из условия приблизительного равенства параметра инерционности $\theta = F_0$ частоте основного тона $F_0 \approx 130$ Гц в речи контрольного диктора [15]. Для иллюстрации сказанного на рис. 6 представлены графики двух СПМ речевого сигнала на выходе выравнивающего фильтра (14), реализованного в двух вариантах оценки СПМ (см. рис. 4). Их огибающие, рассчитанные согласно (16), показаны на рисунке штриховыми кривыми. Даже визуально — по форме спектральной огибающей — метод-прототип проигрывает новому методу АР-моделирования. Полученные результаты отражены на рис. 7 в виде двух гистограмм показателя точности АР-моделирования (15) в зависимости от номера итерации l . Штриховкой здесь отмечены значения показателя для двух представленных на рис. 4 оценок СПМ. Как

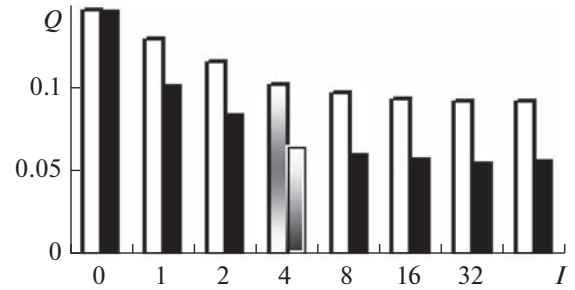


Рис. 7. Гистограммы распределения показателя точности АР-модели (1) гласного звука речи “а” в зависимости от числа итераций (2) при применении двух методов моделирования: с использованием в качестве опорного спектрального образца периодограммы Шустера (светлые столбики) и ее огибающей (черные столбики).

видим, и в отношении точности АР-модели (1) метод, предложенный в данной работе, выигрывает, причем существенно, у своего прототипа: примерно 36.5%. По сравнению с методом Берга (при $l = 0$ на рис. 7) выигрыш возрастает до 56%. Объяснением данного эффекта может служить сама идея АР-моделирования речевого сигнала на основе огибающей его линейчатого спектра мощности.

ЗАКЛЮЧЕНИЕ

В результате проведенного исследования предложен новый метод АР-моделирования речевого сигнала в режиме скользящего окна наблюдений длительностью в один речевой фрейм. Его основополагающая идея состоит в использовании методологии ДСМ с огибающей периодограммы Шустера в качестве опорного спектрального образца. В результате удалось не только решить проблему устойчивости формируемой АР-модели речевого сигнала, но и существенно повысить ее точность в теоретико-информационном смысле.

Полученные результаты предназначены для использования в системах АОР различного назначения, включая речевую связь, прикладную акустику, техническую и медицинскую диагностику, анализ эмоционального состояния дикторов по голосу и др. [2, 20].

СПИСОК ЛИТЕРАТУРЫ

1. Gibson J. // Entropy. 2018. V. 20. № 10. P. 7502018. <https://doi.org/10.3390/e20100750>
2. Gudnason J. Speech Production Modeling and Analysis. Academic Press Library. In Signal Processing, Elsevier. 2014. V. 4. P. 985. <https://doi.org/10.1016/B978-0-12-396501-1.00034-0>
3. Ando Sh. // The J. Acoustical Society of America. 2019. V. 146. P. 2846. <https://doi.org/10.1121/1.5136873>

4. Cui S., Li E., Kang X. // IEEE Int. Conf. Multimedia and Expo (ICME). London. 06–10 Jul. 2020. N.Y.: IEEE, 2020. P. 9102765.
<https://doi.org/10.1109/ICME46284.2020.9102765>
5. Savchenko V.V. // Radioelectronics and Communications Systems. 2021. V. 64. № 11. P. 592.
<https://doi.org/10.3103/S0735272721110030>
6. Castanié F. Digital Spectral Analysis. Parametric, Non-Parametric and Advanced Methods. Hoboken—London: Wiley-ISTE. 2011.
<https://doi.org/10.1002/9781118601877>
7. Rabiner L.R., Shafer R.W. Theory and Applications of Digital Speech Processing. Boston: Pearson, 2010.
8. Marple Jr. S.L. Digital Spectral Analysis with Applications. Mineola, N.Y.: Dover Publications, 2019.
9. Савченко В.В., Савченко Л.В. // РЭ. 2021. Т. 66. № 11. С. 1100.
<https://doi.org/10.31857/S0033849421110085>
10. Kazemipour A., Miran S., Pal P. et al. // IEEE Trans. 2017. V. SP-65. № 9. P. 2333.
<https://doi.org/10.1109/TSP.2017.2656848>
11. Гоноровский И.С. Радиотехнические цепи и сигналы. М.: Сов. радио, 1977.
12. Mustiere F., Bouchard M., Bolic M. // IEEE Trans. 2012. V. ASLP-20. № 2. P. 705.
<https://doi.org/10.1109/TASL.2011.2163511>
13. Savchenko A.V., Savchenko V.V. // Radioelectronics and Communications Systems. 2021. V. 64. № 6. P. 300.
<https://doi.org/10.3103/S0735272721060030>
14. Tohyama M. // Acoustic Signals and Hearing. Acad. Press, 2020. P. 89.
<https://doi.org/10.1016/B978-0-12-816391-7.00013-9>
15. Савченко А.В., Савченко В.В. // Измерит. техника. 2022. № 6. С. 60.
<https://doi.org/10.32446/0368-1025it.2022-6-60-66>
16. Palaparthi A., Titze I.R. // Speech Commun. 2020. V. 123. P. 98.
<https://doi.org/10.1016/j.specom.2020.07.003>
17. Ding J., Tarokh V., Yang Y. // IEEE Trans. 2018. V. IT-64. № 6. P. 4024.
<https://doi.org/10.1109/TIT.2017.2717599>
18. Min S.Y., Kim Y.K. // J. Korea Academia-industrial Cooperation Society. 2010. № 11. P. 3558.
<https://doi.org/10.5762/KAIS.2010.11.9.3558>
19. Савченко В.В. // Научные ведомости Белгород. ГУ. Сер. Экономика. Информатика. 2015. № 7. Вып. 34/1. С. 84.
20. Sharma G., Umapathy K., Krishnan S. // Appl. Acoustics. 2020. V. 158. P. 107020.
<https://doi.org/10.1016/j.apacoust.2019.107020>