

Научная статья

УДК 53.082.4; 004.934

<https://doi.org/10.32446/0368-1025it.2022-10-58-63>

Совершенствование методики измерения показателя точности авторегрессионной модели речевого сигнала

Владимир Васильевич Савченко

Национальный исследовательский университет «Высшая школа экономики», Н. Новгород, Россия, vsavchenko@yandex.ru,
<https://orcid.org/0000-0003-3045-3337>

Аннотация. Рассмотрена задача определения точности авторегрессионной модели речевого сигнала и предложен метод измерений показателя точности в режиме скользящего окна наблюдений. В качестве показателя точности использовано модифицированное значение COSH-расстояния (hyperbolic cosine – функции гиперболического косинуса) авторегрессионной модели относительно одноимённой (однофонемной) периодограммы Шустера в роли опорного спектрального образца. Для изучения возможностей предложенного метода поставлен и проведён натурный эксперимент, в котором объектом исследования служил набор авторегрессионных моделей разных порядков. Указанные модели получены методом Берга для гласных звуков речи контрольного диктора. По результатам выполненных измерений для каждого гласного звука найдены оптимальные значения порядка авторегрессии и соответствующая ему оптимальная авторегрессионная модель. Показано, что данная оптимизация позволила повысить точность авторегрессионной модели речевого сигнала более чем на 60 % в зависимости от звука речи контрольного диктора и особенностей его голосового тракта. Полученные результаты предназначены для использования в системах автоматической обработки и цифровой передачи речи с радикальным сжатием данных на основе коэффициентов линейного предсказания.

Ключевые слова: акустические измерения, акустика речи, голосовой тракт, речевой сигнал, авторегрессионная модель, проблема малых выборок

Для цитирования: Савченко В. В. Совершенствование методики измерения показателя точности авторегрессионной модели речевого сигнала // Измерительная техника. 2022. № 10. С. 58–63. <https://doi.org/10.32446/0368-1025it.2022-10-58-63>

Improving the method for measuring the accuracy indicator of the speech signal autoregressive model

Vladimir V. Savchenko

National Research University Higher School of Economics, N. Novgorod, Russia, vsavchenko@yandex.ru,
<https://orcid.org/0000-0003-3045-3337>

Abstract. The problem of determining the accuracy of an autoregressive model of a speech signal is considered, and a method for measuring the accuracy index in the sliding observation window mode is proposed. As an indicator of the accuracy of the autoregressive model, a modified value of the COSH-distance (functions of the hyperbolic cosine) relative to the eponymous (one-phoneme) Schuster periodogram was used as a reference spectral sample. To study the possibilities of the proposed method, a full-scale experiment was set up and carried out, in which the object of study was a set of autoregressive models of different orders. These models were obtained by Berg's method for the vowel sounds of the controlled speaker's speech. According to the results of the measurements for each vowel, the optimal values of the autoregressive order and the corresponding optimal autoregressive model were found. It is shown that this optimization made it possible to increase the accuracy of the autoregressive model of the speech signal by more than 60 %, depending on the sound of the controlled speaker's speech and the characteristics of his vocal tract. The results obtained are intended for use in automatic processing and digital speech transmission systems with radical data compression based on linear prediction coefficients.

Keywords: acoustic measurements, speech acoustics, voice tract, speech signal, autoregressive model, small sample problem

Введение. Авторегрессионная модель речевого сигнала широко применяется в системах автоматической обработки и передачи речи по цифровым каналам связи в целях кодирования речевой информации с радикальным сжатием данных [1, 2]. В условиях априорной неопределённости тонкой структуры речевого сигнала авторегрессионную модель к нему адаптируют в режиме скользящего окна наблюдений [3, 4].

При этом длина τ окна строго ограничена сверху длительностью интервалов стационарности речевого сигнала [5]. Как следствие, на практике часто возникает проблема малых выборок [6, 7]. Она проявляется в недостаточной точности используемой авторегрессионной модели, в том числе, связанной с необоснованным выбором порядка $p < \infty$ [8]. Острота указанной проблемы становится особенно

очевидной, если учесть, что в теории отсутствует строгий критерий точности авторегрессионной модели, а существующие критерии, например Акаике, Шварца и др. [9], мало пригодны для применения к речевому сигналу, так как рассчитаны на однородные эргодические процессы. Поэтому актуальной является задача измерения показателя точности авторегрессионной модели по выборке конечного объема $N = \tau F$, где F – частота дискретизации наблюдений.

Цель исследования – разработка методологических основ оптимизации авторегрессионных моделей речевого сигнала с учётом его вариативности и особенности голосового тракта диктора [10, 11]. Статья написана в развитие результатов предыдущих работ [2, 3].

Постановка задачи. Поскольку авторегрессионная модель непосредственно связана с огибающей линейчатого спектра мощности речевого сигнала [12, 13], то определим этот спектр в пределах фрейма речи $x(n)$, где $n=0, 1, \dots, N-1$. При выборе способа спектрального анализа будем следовать логике работ [7, 14] и отдадим предпочтение дискретному фурье-преобразованию как разновидности линейной обработки сигналов, которая не связана с известными эффектами параметрических оценок спектра мощности: смещения и расщепления спектральных линий в условиях малых выборок наблюдений. Основанная на дискретном фурье-преобразовании периодограмма Шустера [15]:

$$G_x(f_m) = (FM)^{-1} \left| \sum_{n=0}^{M-1} x(n) \exp(-j2\pi n m M^{-1}) \right|^2, \quad (1)$$

$$m = 0, 1, \dots, M-1$$

определяет мгновенную оценку спектра мощности (ОСМ) сигнала на выделенном наборе дискретных частот $f_m = m \Delta f = m M^{-1} F$. При равенстве $M = 2^k > N$, где k – некоторое целое число, периодограмма рассчитана на применение алгоритмов быстрого преобразования Фурье [16]. При этом только первые N из M отсчётов временного ряда $\{x(n)\}$ в правой части (1) отличны от нуля. Например, при длительности речевого фрейма $\tau = 20 \dots 30$ мс и частоте дискретизации $F = 8$ кГц (согласована с полосой пропускания стандартного телефонного канала связи) для $k = 10$ размерность составит $M = 1024 \gg N = 160 \dots 240$.

Согласно методологии параметрического спектрального анализа определим авторегрессионную модель в частотной области выражением общего вида [17]:

$$G_p(f_m) = \sigma_p^2 T \left| 1 - \sum_{i=1}^p a_p(i) \exp(-j2\pi i m M^{-1}) \right|^{-2}, \quad (2)$$

где $T = F^{-1}$; σ_p^2 – масштабный множитель; $a_p(i)$ – i -й элемент вектора коэффициентов линейной авторегрессии p -го порядка. Параметры σ_p^2 и $a_p(i)$ определяются по выборке $\{x(n)\}$ в соответствии с используемым методом спектрального анализа. Например, в методе Берга в указанных целях рекуррентная вычислительная процедура вида [15]:

$$\forall q = \overline{1, p}: a_q(i) = a_{q-1}(i) + c_q a_{q-1}(i-1), \quad i = \overline{1, q}; \quad (3)$$

$$c_q = \frac{2 \sum_{n=q}^{N-1} \eta_{q-1}(n) v_{q-1}(n-1)}{\sum_{n=q}^{N-1} [\eta_{q-1}^2(n) + v_{q-1}^2(n-1)]};$$

$$\eta_q(n) = \eta_{q-1}(n) + c_q v_{q-1}(n-1);$$

$$v_q(n) = v_{q-1}(n-1) + c_q \eta_{q-1}(n), \quad n = 0, 1, \dots, N-1,$$

применяется при её инициализации системой равенств

$$\forall n < N: v_0(n) = \eta_0(n-1) = x(n).$$

Финальные значения рекурсии (3) совместно с (2) определяют авторегрессионную модель спектральной огибающей $\bar{G}_x(f_m)$ речевого сигнала. Точность этой модели зависит от её порядка. Масштабный множитель σ_p^2 здесь роли не играет, поскольку никак не связан с формой спектральной огибающей. Поэтому задача состоит в определении показателя точности авторегрессионной модели (2) в зависимости от порядка p и его последующем измерении в режиме скользящего окна наблюдений $\{x(n)\}$.

Показатель точности. Перепишем выражение (2) в виде инверсного преобразования

$$G_p(f_m) = \sigma_p^2 T |A_p(jf_m)|^{-2}$$

квадрата модуля комплексного коэффициента передачи $A_p(jf_m)$ трансверсального фильтра p -го порядка

$$A_p(jf_m) = 1 - \sum_{i=1}^p a_p(i) \exp(-j2\pi i m M^{-1}), \quad m = \overline{0; M-1}. \quad (4)$$

Этот фильтр используем для выравнивания в частотной области огибающей $\bar{G}_x(f_m)$ периодограммы Шустера (1), которая выполняет в данном случае роль опорного спектрального образца [3]. Сигнал на выходе выравнивающего фильтра

$$G_y(f_m; p) = |A_p(jf_m)|^2 G_x(f_m) = \sigma_p^2 T \frac{G_x(f_m)}{G_p(f_m)}, \quad m = \overline{0; M-1}, \quad (5)$$

определяет ОСМ последовательности $\{y(n)\}$ импульсов возбуждения голосового тракта условного диктора в рамках обратной задачи речеобразования [18]: от речевого сигнала – к его голосовому источнику.

В идеале ОСМ (5) представляет собой прямоугольную последовательность $K = 0,5 F F_0^{-1}$ гармонических составляющих со сдвигом между собой на частоту основного тона F_0 [19]. Однако на практике форма спектральной огибающей $\bar{G}_y(f_m; p)$ может существенно отличаться от прямоугольной, прежде всего из-за неидеальности используемой авторегрессионной модели (2). С указанной точки зрения принципиальное значение имеет порядок модели. Оптимизируем значение p в пределах конечного множества (R -множества) альтернатив $p_1 < p_2 < \dots < p_R$.

В основу оптимизации положим очевидное соображение: чем выше точность авторегрессионной модели, тем более ярко выражены в ОСМ (5) гармонические составляющие и тем более близка форма их огибающей $\{\overline{G}_Y(f_m; p)\}$ к прямоугольной $G_0 = \text{const}$ для всех $m < M$. При учёте сказанного воспользуемся в качестве критериальной функции модифицированной мерой COSH-расстояния (COSH, hyperbolic cosine – функции гиперболического косинуса) $\rho(p)$ со свойством масштабной инвариантности в частотной области [20]:

$$\begin{aligned} \rho(p) &= \sqrt{\left[M^{-1} \sum_{m=0}^{M-1} G_0 \bar{G}_y^{-1}(f_m; p) \right] \left[M^{-1} \sum_{m=0}^{M-1} \bar{G}_y(f_m; p) G_0^{-1} \right]} - 1 = \\ &= M^{-1} \sqrt{\left[\sum_{m=0}^{M-1} \bar{G}_y^{-1}(f_m; p) \right] \left[\sum_{m=0}^{M-1} \bar{G}_y(f_m; p) \right]} - 1 \geq 0. \end{aligned} \quad (6)$$

Мера (6) тождественно равна нулю при равенстве спектральной огибающей произвольной константе: $G_Y(f_m; p) = G_0 \quad \forall m < M$. Как следует из результатов работы [3], именно в таком варианте ОСМ (5) точность авторегрессионной модели (2) будет максимальной. С увеличением $\rho(p)$ точность уменьшается. Мету (6) можно рассматривать в таком случае в роли объективного показателя точности формируемой авторегрессионной модели. Отсюда вытекает следующий критерий принятия решений в задаче оптимизации порядка модели на множестве R рассматриваемых альтернатив:

$$p^* = \underset{\forall r \leq R}{\text{Arg min}} \rho(p) \Big|_{p=p_r}. \quad (7)$$

Оптимальная модель отвечает выражению (2) при выборе её порядка $p \leq p_R$ согласно правилу (7). Практическая реализация сводится к измерению показателя (6) в режиме скользящего окна наблюдений.

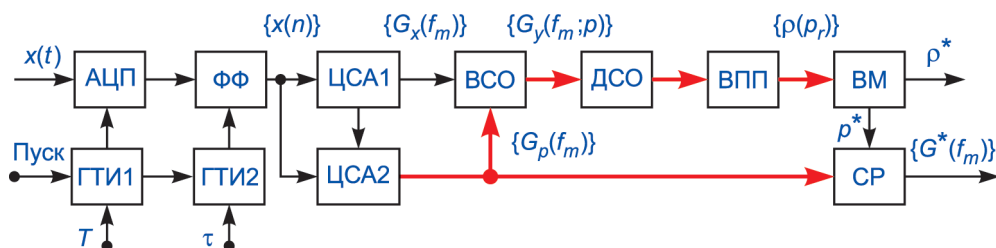


Рис. 1. Структурная схема устройства измерения показателя точности и оптимизации порядка авторегрессионной модели речевого сигнала:

АЦП – аналого-цифровой преобразователь; ФФ – формирователь фреймов; ЦСА1, ЦСА2 – блоки цифрового спектрального анализа; ВСО – блок выравнивания спектральной огибающей; ДСО – детектор спектральной огибающей; ВПП – блок вычисления парциальных показателей; ВМ – схема выбора минимума; СР – блок считывания результата; ГТИ1, ГТИ2 – генераторы тактовых импульсов

Fig. 1. Block diagram of the device for measuring the quality indicator of the AR modeling of a speech signal:

ADC – analog-to-digital converter; FB – frame builder; DSA1, DSA2 – the digital spectral analysis blocks; SEF – spectral envelope flattening block; ED – envelope detector; BCPI – block for calculating partial indicators; MSS – minimum selection scheme; RR – result reading block; CPG1, CPG2 – the clock pulses generators

Метод измерений показателя точности. Структурная схема устройства для измерений показателя точности (6) и оптимизации порядка авторегрессионной модели (2) представлена на рис. 1. Устройство содержит последовательно соединённые элементы: аналого-цифровой преобразователь *АЦП*; формирователь фреймов *ФФ*; первый блок цифрового спектрального анализа *ЦСА1*; блоки выравнивания *ВСО* и детектирования *ДСО* спектральной огибающей; блоки вычисления *ВПП* и выбора минимума *ВМ* парциальных показателей точности, определения и считывания результата *СР*; два генератора тактовых импульсов *ГТИ1*, *ГТИ2*; второй блок *ЦСА2*. Красными линиями обозначены многомерные (размерностью *M*) функциональные связи. Регулируемые параметры устройства – период дискретизации *T* речевого сигнала и длина τ окна наблюдений.

Работа устройства начинается с запуска генераторов, необходимых для синхронизации работы отдельных элементов вычислительного процесса (3)–(6). С помощью генераторов периодически (с периодом τ) формируются и обновляются фреймы речевого сигнала $\{x(n)\}$ на входах блоков ЦСА1, ЦСА2. Соответственно, периодически будут обновляться ОСМ (1), (2) на их выходах. Причём согласно (2) модель обновляется одновременно в R разных вариантах $\{G_p(f_m)\}$ в зависимости от порядка $p=p_r$ для всех $r \leq R$, т. е. в соответствии с выражением (5) будет R вариантов спектральной огибающей $\underline{G}_y(f_m; p)$. Для их выделения в устройстве служит детектор спектральной огибающей. Его практическая реализация по схеме рециркулятора [16]:

$$\overline{G}_Y(f_m; p) = b\overline{G}_Y(f_m - \Delta f; p) + G_Y(f_m; p), \quad m = \overline{1, M}, \quad (8)$$

основана на процедуре накопления или сглаживания речевого сигнала в частотной области. Константа b в правой части (8) влияет на инерционность такого накопления. Количественный показатель $\theta = -\Delta f / \ln b$ инерционности должен быть согласован с частотой F_0 основного тона речевого сигнала. Отсюда в первом приближении запишем

$$\{p(p_i)\}$$

$$b = \exp(-\Delta f/F_0) = \exp(-F/(MF_0))$$

и, например, в диапазоне рабочих значений $F_0=100\ldots 200$ Гц [3] при $F=8$ кГц, $M=1024$ получим $b=0.93\ldots 0.97$.

Основной результат работы устройства – показатель потенциально достижимой точности авторегрессионной модели (2) – формируется в R -канальной схеме блока BM (см. рис. 1) по критерию (7) выбора оптимального порядка. В соответствии с порядком p^* с помощью блока CP считывается оптимальная авторегрессионная модель $\{G^*(f_m)\}$.

Программа и результаты эксперимента. В эксперименте использован пилотный вариант устройства (см. рис. 1), реализованный на базе авторского программного обеспечения Phoneme Training¹. Программа размещена в открытом доступе, её интерфейс подробно описан в работе [11].

Объектом экспериментального исследования выбраны сигналы шести гласных звуков русской речи контрольного диктора – автора настоящей статьи. Достаточно большая (2–3 с) длительность каждого сигнала изначально рассчитана на его автоматическое последовательное членение на множество фреймов длительностью $\tau=30$ мс со сдвигом по времени на 10 мс. При частоте дискретизации $F=8$ кГц размерность фрейма N составила 240 отсчётов. Форма окна наблюдений – прямоугольная. В результате для каждого гласного звука речи сформирована представительная база данных объёмом $L=200\dots300$ одноимённых образцов. Далее по каждому образцу рассчитана периодограмма Шустера (1) размерностью $M=1024$ и с шагом по частоте $\Delta f=8000:1024=7,8125$ Гц. Параллельно в качестве авторегрессионной модели речевого сигнала определены спектральные оценки Берга (2) в $R=40$ альтернативных вариантах порядка $p=10\dots50$. В соответствии с рекурсией Левинсона (3) суммарная сложность вычислительной процедуры составила при этом порядка $3Np_R=3\cdot240\cdot50=36000$ элементарных операций умножения-сложения на интервале один речевой фрейм. Этот объём с запасом отвечает производительности современных речевых систем и технологий [21].

В процессе эксперимента для каждой из 40 альтернатив при $b=0,96$ согласно (8) получены оценки спектральной огибающей $\bar{G}_y(f_m; p)$, по которым в соответствии с критерием (7) были установлены оптимальное значение порядка p^* , соответствующая авторегрессионная модель (2) и показатель (6) её точности. Затем результаты усреднены по результатам L испытаний для каждой гласной фонемой контрольного диктора. В итоге относительная погрешность экспериментальных оценок [3] $\varepsilon=1,65L^{-1/2}=1,65(200\dots300)^{-1/2}$ при доверительной вероятности 0,9 составила не более 10–12 %.

На рис. 2 показан график рабочей характеристики $\rho(p)$ предложенного метода (1)–(7) применительно к гласному звуку речи «а». Вертикальные отрезки в контрольных точках данного графика определяют границы доверительного интервала значений показателя по результатам многократных измерений. Кривая $\rho(p)$ имеет единственный минимум $p^*=32$, что свидетельствует об эффективности критерия (7) и в целом метода акустических измерений. Достижимый эффект может быть охарактеризован значением выигрыша B по точности авторегрессионной модели (2) за счёт оптимизации её порядка p^* по отношению к его стандартному значению $p=10$ [2]:

$$B = \frac{\rho(10) - \rho^*}{\rho(10)} \cdot 100. \quad (9)$$

Согласно данным рис. 2 рассчитанное по формуле (9) значение составило $B=79,48\%$.

Для иллюстрации полученного результата на рис. 3 представлено семейство спектральных оценок Берга для авторегрессионных моделей порядков $p=10; 32; 50$ (кривые 1–3) в сравнении с периодограммой Шустера (кривая 4) на примере одного и того же (типичного) фрейма $\{x(n)\}$ гласного звука «а». Все кривые нормированы по средней мощности к единице. Оптимальной по точности является модель порядка $p=32$. На рис. 4 приведён соответствующий порядку $p=32$ график ОСМ (5) на выходе выравнивающего фильтра (4), а также огибающая $\bar{G}_y(f_m; p)$ (показана штриховой линией). Её форма близка к прямоугольной. Показатель (6) в данном случае характеризуется минимумом $\rho^*=\rho(32)\approx 0,04$ (см. рис. 2).

Сделанные выводы сохраняют справедливость и для остальных гласных звуков речи контрольного диктора. При этом их оптимальные по критерию (7) значения порядка модели могут сильно различаться между собой, например,

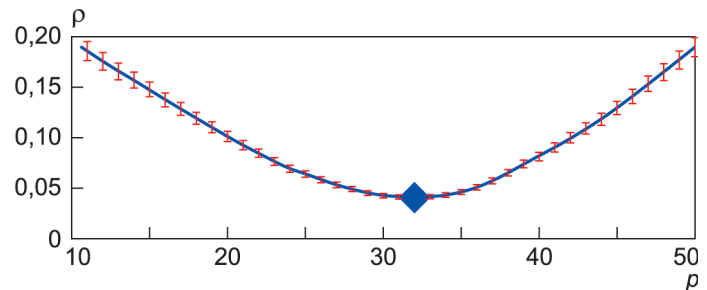


Рис. 2. Рабочая характеристика (с границами доверительного интервала по результатам многократных измерений показателя точности) предложенного метода измерения точности авторегрессионной модели на примере гласного звука «а» речи контрольного диктора

Fig. 2. The operating characteristic of the proposed method on the example of the vowel sound of speech “a” from the control speaker

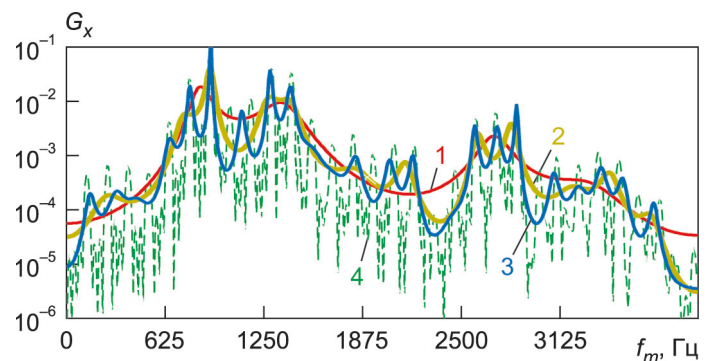


Рис. 3. Спектральные оценки Берга для авторегрессионных моделей порядков $p=10; 32; 50$ (кривые 1–3) гласного звука «а» в сопоставлении с периодограммой Шустера (кривая 4)

Fig. 3. Berg spectral estimates for the speech sound frame “a” of orders $p=10; 32; 50$ (curves 1–3) against the background of the Schuster periodogram (curve 4)

¹ Phoneme Training. Информационная система фонетического анализа и обучения речи [сайт]. URL: <https://sites.google.com/site/frompldcreators/produkty-1/phonemetraining> (дата обращения: 20.09.2022).

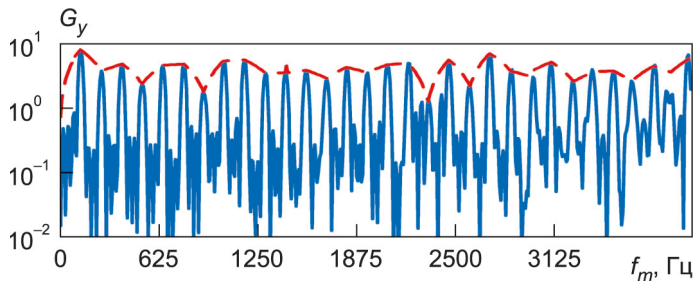


Рис. 4. Графики нормированной оценки спектра мощности (5) сигнала звука речи «а» на выходе выравнивающего фильтра (4) при порядке $p=32$ авторегрессионной модели

Fig. 4. Graphs of the normalized power spectrum estimation (5) of the speech sound signal "a" at the output of the flattening filter (4) when its order is $p=32$

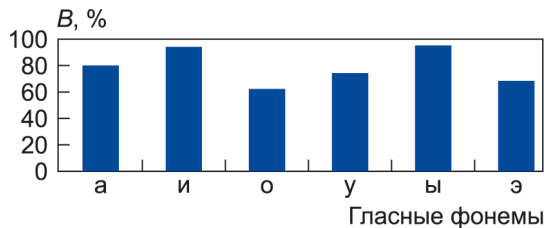


Рис. 5. Гистограмма распределения выигрыша B (9) по сигналам гласных фонем

Fig. 5. Histogram of distribution of payoff (9) by vowel phoneme signals

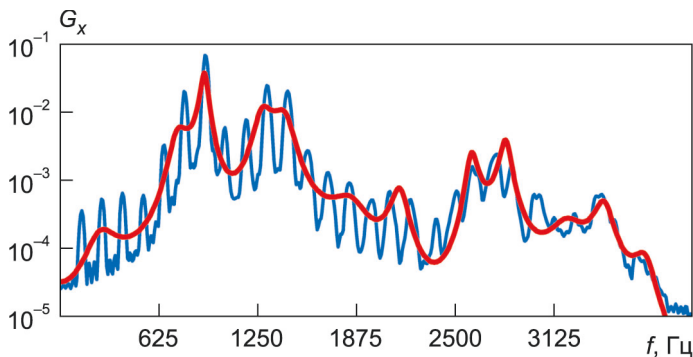


Рис. 6. Усреднённая периодограмма Шустера (синяя линия) в сравнении с оценкой Берга порядка $p=32$ (красная линия) на входе выравнивающего фильтра (4)

Fig. 6. The average Schuster periodogram (blue line) compared with the Berg estimate of the order $p=32$ (red line) at the input of the flattening filter (4)

для звука «о» имеем $p^*=22$. Различаются при этом и достигаемые в каждом случае значения выигрыша (9). Гистограмма их распределения по гласным фонемам представлена на рис. 5. Оптимизация позволила повысить точность авторегрессионной модели речевого сигнала более чем на 60 % в зависимости от звука речи контрольного диктора и особенностей его голосового тракта. Максимальный выигрыш достигнут для гласных звуков верхнего подъёма «и», «у», «ы», что можно отнести к особенности конкретного диктора.

Обсуждение полученных результатов. В отличие от метода Берга периодограмма Шустера в настоящее время почти не находит самостоятельного применения в задачах спектрального анализа временных рядов ввиду её не вполне подходящих статистических характеристик [15]. Вместо периодограммы Шустера в этих целях применяют методы Бартлетта, Уэлча и другие, основанные на эффекте усреднения периодограммных ОСМ по выборкам наблюдений достаточно большого объёма $N \gg M$. Однако по условиям рассматриваемой в данной статье задачи выполняется прямо противоположное соотношение $M \gg N$, которое исключает статистическое усреднение и этим ограничивает возможности периодограммы Шустера в роли ОСМ речевого сигнала. Поэтому возникает закономерный вопрос в отношении обоснованности выражения (1) в качестве опорного спектрального образца. Ответом на него служат результаты второго, заключительного, этапа экспериментальных исследований, в рамках которого была изучена сравнительная эффективность предложенного выше метода в двух вариантах задания спектрального образца: в практически реализуемом варианте периодограммы (1), а также в умозрительном (гипотетическом) варианте Уэлча

$$G_x(f_m) = I^{-1} \sum_{i=1}^I G_x^{(i)}(f_m), \quad \forall m \leq M-1, \quad (10)$$

с I -кратным усреднением парциальных периодограмм Шустера на интервале наблюдений суммарной длительностью t/I . Здесь символом $G_x^{(i)}(f_m)$ обозначена парциальная периодограмма (1) в границах i -го речевого фрейма $\{x(n)\}$ конечного объёма N .

Полученные результаты представлены на рис. 6 в виде графиков двух спектральных оценок сигнала фонемы «а» контрольного диктора: Уэлча (10) при $I=240$ (синяя линия) и Берга (красная линия) при оптимальном порядке $p^*=32$ модели (2). Оценка Берга не усреднялась, т. е. использовался один из фреймов $\{x(n)\}$, тем не менее она хорошо согласуется с огибающей усреднённой периодограммной оценки.

Огибающая ОСМ на выходе выравнивающего фильтра в данном случае не является более равномерной по сравнению с аналогичным графиком на рис. 4. Более того, показатель (6) её точности $\rho(32)=0,051$ не вышел за пределы соответствующего доверительного интервала $[0,02; 0,06]$ в точке оптимума рабочей характеристики (см. рис. 2). Отсюда можно сделать вывод, что усреднение периодограммы Шустера по методу Уэлча (10) почти ничего не даёт с точки зрения повышения точности формируемой авторегрессионной модели (2). Периодограмма Шустера в данном случае играет роль не итоговой ОСМ, а источника полезной информации в отношении тонкой структуры речевого сигнала. В этом качестве она безупречна в силу линейности обработки речевого сигнала при дискретном преобразовании Фурье [7].

Заключение. В результате проведённого исследования предложены показатель точности авторегрессионной модели речевого сигнала и метод его измерения в режиме скользящего окна наблюдений конечной длительности t . В рамках предпринятой оптимизации авторегрессионной

модели спектральная оценка Берга и периодограмма Шустера взаимно дополняют друг друга, хотя на практике они обычно конкурируют между собой [15, 17]. При этом критерий оптимизации впервые строго сформулирован в терминах обратной задачи речеобразования. В отличие от критериев типа AIC, BIC и им подобных [8, 9] предложенный критерий не требует синхронизации окна наблюдений с периодом основного тона речевого сигнала, что сильно ослабляет остроту проблемы априорной неопределённости [3].

Полученные результаты предназначены для использования в системах автоматической обработки и цифровой передачи речи с радикальным сжатием данных на основе коэффициентов линейного предсказания [1, 2].

Список источников / References

1. Gibson J., *Entropy*, 2018, vol. 20, no. 10, 7502018. <https://doi.org/10.3390/e20100750>
2. Savchenko V. V., *Radioelectronics and Communications Systems*, 2021, vol. 64, no. 11, pp. 592–603. <https://doi.org/10.3103/S0735272721110030>
3. Савченко А. В., Савченко В. В. Адаптивный метод измерения частоты основного тона с использованием двухуровневой авторегрессионной модели речевого сигнала // *Измерительная техника*. 2022. № 6. С. 60–66. <https://doi.org/10.32446/0368-1025it.2022-6-60-66> [Savchenko V. V., Savchenko A. V., An adaptive method for measuring the pitch frequency using a two-level autoregressive model of a speech signal, *Izmeritel'naya tekhnika*, 2022, no. 6, pp. 60–66. (In Russ.)].
4. Jaramillo E., Nielsen J. K., Christensen M. G., *27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5. <https://doi.org/10.23919/EUSIPCO.2019.8902763>
5. Савченко В. В. Исследование стационарности случайных временных рядов с использованием принципа минимума информационного рассогласования // *Известия вузов. Радиофизика*. 2017. Т. 60. № 1. С. 89–96. [Savchenko V. V., *Radio-physics and Quantum Electronics*, 2017, vol. 60, no. 1, pp. 89–96. <https://doi.org/10.1007/s11141-017-9778-y>].
6. Cui S., Li E., Kang X., *IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, 2020, pp. 1–6. <https://doi.org/10.1109/ICME46284.2020.9102765>
7. Ando Sh., *The Journal of the Acoustical Society of America*, 2019, vol. 146, no. 11, 2846. <https://doi.org/10.1121/1.5136873>
8. Savchenko V. V., Savchenko A. V., *Radioelectronics and Communications Systems*, 2019, vol. 62, no. 5, pp. 276–286. <https://doi.org/10.3103/S0735272719050042>

9. Ding J., Tarokh V., Yang Y., *IEEE Transactions on Information Theory*, 2018, vol. 64, no. 6, pp. 4024–4043. <https://doi.org/10.1109/TIT.2017.2717599>

10. Savchenko V. V., *Radioelectronics and Communications Systems*, 2020, vol. 63, pp. 532–542. <https://doi.org/10.3103/S0735272720100039>

11. Савченко В. В., Савченко Л. В. Метод измерений показателя разборчивости речевого сигнала в информационной метрике Кульбака-Лейблера // *Измерительная техника*. 2019. № 9. С. 59–64.

<https://doi.org/10.32446/0368-1025it.2019-9-59-64> [Savchenko V. V., Savchenko L. V., *Measurement Techniques*, 2019, vol. 62, no. 9, pp. 832–839. <https://doi.org/10.1007/s11018-019-01702-1>].

12. Tohyama M., Spectral envelope and source signature analysis, in: *Acoustic Signals and Hearing*, Academic Press, 2020, pp. 89–110. <https://doi.org/10.1016/B978-0-12-816391-7.00013-9>

13. Sun P., Mahdi A., Xu J., Qin J., *Speech Communication*, 2018, vol. 101, pp. 57–69. <https://doi.org/10.1016/j.specom.2018.05.006>

14. Савченко В. В., Савченко Л. В. Метод авторегрессионного моделирования речевого сигнала на основе его дискретного фурье-преобразования и масштабно-инвариантной меры информационного рассогласования // *Радиотехника и электроника*. 2021. Т. 66. № 11. С. 1100–1108. <https://doi.org/10.31857/S0033849421110085> [Savchenko V. V., Savchenko L. V., *Journal of Communications Technology and Electronics*, 2021, vol. 66, no. 11, pp. 1266–1273. <https://doi.org/10.1134/s1064226921110085>].

15. Marple S. L., *Digital spectral analysis with applications*, 2nd ed., Mineola, New York, Dover Publications, 2019, 432 p.

16. Радиоэлектронные системы. Основы построения и теория: Справочник / Под ред. Я. Д. Ширмана. Изд. 2-е, перераб. и доп. М.: Радиотехника, 2007, 657 с. [Radioelektronnyye sistemy. Osnovy postroeniya i teoriya: Spravochnik, Ed. Ya. D. Shirman, 2nd ed., Moscow, Radiotekhnika Publ., 2007, 657 c. (In Russ.)].

17. Benesty J., Chen J., Huang Y., Linear Prediction, in: *Springer handbook of speech processing, Part B*, N.Y., Springer, 2008, pp. 111–124. https://doi.org/10.1007/978-3-540-49127-9_7

18. Palaparthi A., Titze I. R., *Speech Communication*, 2020, vol. 123, pp. 98–108. <https://doi.org/10.1016/j.specom.2020.07.003>

19. Rabiner L. R., Shafer R. W., *Theory and Applications of Digital Speech Processing*, Boston, Pearson, 2010, 1060 p.

20. Savchenko A. V., Savchenko V. V., Savchenko L. V., *Optimization Letters*, 2021, no. 7. <https://doi.org/10.1007/s11590-021-01790-5>

21. Kumar S., Singh S. K., Bhattacharya S., *International Journal of Speech Technology*, 2015, vol. 18, pp. 521–527. <https://doi.org/10.1007/s10772-015-9296-2>

Статья поступила в редакцию 30.06.2022; одобрена после рецензирования 06.09.2022; принята к публикации 08.09.2022. The article was submitted 30.06.2022; approved after reviewing 06.09.2022; accepted for publication 08.09.2022.