

МЕТОД КОРРЕКТИРОВКИ КОЭФФИЦИЕНТОВ ЛИНЕЙНОГО ПРЕДСКАЗАНИЯ ДЛЯ СИСТЕМ ЦИФРОВОЙ ОБРАБОТКИ РЕЧИ СО СЖАТИЕМ ДАННЫХ НА ОСНОВЕ АВТОРЕГРЕССИОННОЙ МОДЕЛИ ГОЛОСОВОГО СИГНАЛА

© 2024 г. В. В. Савченко^а, *, Л. В. Савченко^б

^аРедакция журнала “Радиотехника и электроника”,
ул. Моховая, 11, корп. 7, Москва, 125009 Российская Федерация

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

*E-mail: vvsavchenko@yandex.ru

Поступила в редакцию 18.09.2023 г.

После доработки 18.12.2023 г.

Принята к публикации 16.01.2024 г.

Рассмотрена проблема искажений авторегрессионной модели голосового сигнала под действием аддитивного фонового шума в системах цифровой обработки речи со сжатием данных на основе линейного предсказания. В частотной области указанные искажения проявляются в ослаблении основных формант, отвечающих за разборчивость речи диктора. Для компенсации формантного ослабления предложено корректировать основные параметры авторегрессионной модели — коэффициенты линейного предсказания. Разработан регулярный метод их корректировки с использованием импульсной характеристики рекурсивного формирующего фильтра. При применении данного метода наряду с амплитудным усилением формант их частоты сохраняются неизменными как фактор узнаваемости голоса диктора. Эффективность метода исследована экспериментально с использованием авторского программного обеспечения. По результатам проведенного эксперимента сделаны выводы о существенном повышении относительного уровня формант в спектре мощности откорректированного голосового сигнала.

Ключевые слова: теория сигналов, голосовой сигнал, цифровая обработка речи, цифровая передача речи, спектральный анализ, спектральная плотность мощности, дискретное спектральное моделирование, авторегрессионная модель, all-pole model

DOI: 10.31857/S0033849424040056, EDN: JSCORK

ВВЕДЕНИЕ

На протяжении нескольких десятков лет авторегрессионная модель (АР-модель) находит широкое распространение в системах цифровой обработки речи (ЦОР) со сжатием данных [1–3]. Данная модель положена в основу целого семейства современных алгоритмов кодирования с линейным предсказанием¹, объединенных общей аббревиатурой LPC (linear predictive coding) и широко представленных в серии рекомендаций G Международного союза электросвязи (International telecommunication union,

ITU), таких как G.728, G.729 и др.² От качества используемой в них АР-модели зависит разборчивость речи диктора и узнаваемость тембра его голоса [4–6]. А это два основных требования действующего государственного стандарта к системам речевой связи³.

Качество АР-модели определяется, в свою очередь, точностью задания ее параметров. Ими служат коэффициенты линейного предсказания (КЛП). В процессе обработки речи их статистически оценивают по конечной выборке (фрейму) голосового сигнала [7, 8]. Для этого разработан эффективный математический аппарат в области цифрового спектрального анализа [9]. Предпочтение при этом отдают параметрическим методам спектрального анализа, таким, например, как методы Берга и Юла–Уолкера [10], которые обладают повышенной разрешающей способностью по частоте в условиях ограниченной длительности $\tau = 10...20$ мс стандартного речевого фрейма. АР-модель служит их теоретическим обоснованием и одновременно средством

¹ ГОСТ Р 53556.3–2009. Звуковое вещание цифровое. Кодирование сигналов звукового вещания с сокращением избыточности для передачи по цифровым каналам связи. Часть III (MPEG-4 audio). Кодирование речевых сигналов с использованием линейного предсказания.

² <https://www.itu.int/rec/T-REC-G/en>

³ ГОСТ Р 50840–95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости.

эффективного сжатия данных в системах цифровой передачи речи по низкоскоростным каналам связи⁴. В условиях не слишком напряженной помеховой обстановки такая модель обеспечивает достаточно высокую эффективность ЦОР [4, 7].

Однако при действии интенсивного акустического (фонового) шума на входе точность формируемой АР-модели неизбежно ухудшается, ее параметры искажаются [11]. Чем больше интенсивность шума, тем больше искажаются оценки КЛП. В частотной области указанная проблема проявляется в пропорциональном ослаблении амплитуды формант голосового сигнала, что негативно сказывается на разборчивости речи диктора, переданной по цифровому каналу связи [12, 13]. Поэтому в рассматриваемых условиях параметры АР-модели нуждаются в определенной корректировке [11, 14].

К сожалению, данная задача не имеет оптимального решения. Препятствием принципиального характера в ней служит априорная неопределенность очередного (наблюдаемого в режиме скользящего окна) звука речи диктора [3, 15]. В рамках проведенного в статье исследования данное препятствие преодолевается с использованием авторской методики формантного анализа голосового сигнала во временной области.

Цель исследования — разработка регулярно метода корректировки КЛП для систем ЦОР с кодированием голосового сигнала на основе АР-модели.

1. АР-МОДЕЛЬ ГОЛОСОВОГО СИГНАЛА

В соответствии с акустической теорией речеобразования [1, 4] форманты голосового сигнала определяют через распределение по частоте областей локальных максимумов в спектре мощности анализируемого звука речи (фонемы) от конкретного диктора. Применительно к гласным фонемам каждая форманта характеризуется своей амплитудой и частотой [16]. Их связывают с резонансами голосового тракта диктора [2]. За частоту форманты в таком случае принимают частоту гармоники основного тона с максимальной амплитудой в пределах области соответствующего локального максимума. При этом количество формант зависит от полосы анализируемых частот. Так, в полосе частот стандартного телефонного канала связи обычно насчитывают четыре-пять разных формант. Из них ключевое значение с точки зрения разборчивости речи и узнаваемости голоса диктора имеют две первые форманты, пронумерованные в порядке возрастания своих частот [12]. В отсутствие фонового

шума и явных отклонений в функционале голосового тракта их уровень на 20...30 дБ превышает средний уровень спектра мощности голосового сигнала. Считается [17], что этого вполне достаточно для безошибочного восприятия речи диктора условным слушателем. Ситуация в корне меняется при действии аддитивного фонового шума.

В этом случае на входе системы ЦОР суммируются (накладываются друг на друга) два разных сигнала: звук речи $s(t)$ условного диктора и акустический шум $\eta(t)$. Обозначим суммарный сигнал через

$$x(t) = s(t) + \eta(t). \quad (1)$$

АР-модель речевого сигнала (1) в дискретном времени t определяется линейным разностным уравнением

$$y(t) = \sum_{i=1}^p a_i y(t-i) + \xi(t), \quad t = 0, 1, \dots, \quad (2)$$

динамики рекурсивного фильтра p -го порядка с порождающим белым шумом $\xi(t)$ на входе. Параметрами этого фильтра служат статистические оценки КЛП $\{a_i\}$. Будем называть этот фильтр формирующим [15]. Если связать его сигнал возбуждения с погрешностью моделирования

$$z(t) = x(t) - \sum_{i=1}^p a_i x(t-i), \quad t = 0, 1, \dots,$$

то выражение (2) можно рассматривать в роли внутрипериодного (внутри периода основного тона) предиктора речи [3, 18]. Это ключевое звено системы ЦОР с линейным предсказанием [4]. Его оценивают по выборке наблюдений $\{x(t)\}$ конечного объема $n = \tau/T = \tau F$, где T — период взятия отсчетов при частоте дискретизации голосового сигнала $F = 1/T$. Для стандартного телефонного канала выполняется равенство $F = 8$ кГц.

Под действием аддитивного фонового шума $\eta(t)$ используемые в (2) оценки КЛП неизбежно искажаются. Как результат, искажается и оценка спектра мощности голосового сигнала. Причем характер указанных искажений в значительной мере зависит от спектральных характеристик фонового шума $\eta(t)$. Первостепенный интерес в этом смысле представляет белый шум $\eta(t)$ как случайный процесс с максимальной энтропией [9]. Под его воздействием существенно меняется тонкая структура голосового сигнала. Форманты “размываются” по частоте, их амплитуды уменьшаются, и это отрицательно сказывается на качестве речевой связи [19].

Наглядное представление о существовании рассматриваемой проблемы дает рис. 1, на котором показано семейство графиков АР-оценки внутрипериодного спектра мощности сигнала гласной фонемы “а”, произнесенной контрольным диктором, для трех значений отношения сигнал-шум (ОСШ) q^2 по средней мощности. Они получены экспериментально, согласно выражению [10]

⁴ ГОСТ Р 51061–97. Системы низкоскоростной передачи речи по цифровым каналам связи. Параметры качества речи и методы измерений.

$$G_p(f) = T \left| 1 - \sum_{i=1}^p a_i \exp(-j2\pi i f T) \right|^{-2} = \quad (3)$$

$$= T |K_p(jf)|^2, \quad |f| \leq 0.5F,$$

нормированной (по дисперсии порождающего шума $\xi(t)$) оценки огибающей спектральной плотности мощности (СПМ) голосового сигнала фиксированного порядка⁵ $p = 10$. Здесь символом j обозначена мнимая единица. Введем обозначение [14]

$$K_p(jf) = \left(1 + \sum_{i=1}^p [-a_i \exp(-j2\pi i f T)] \right)^{-1} = \quad (4)$$

$$= \left(1 + \sum_{i=1}^{\infty} b_i \exp(-j2\pi i f T) \right)^{-1} = \frac{T}{\text{DFT}\{\mathbf{b}_p\}}$$

комплексного коэффициента передачи формирующего фильтра через инверсию $1/\text{DFT}(\mathbf{b}_p)$ дискретного фурье-преобразования (ДФП) расширенного вектора коэффициентов $\mathbf{b}_p = \{b_i\} = (1, -a_1, -a_2, \dots, -a_p, 0, 0, \dots) = (1, -\mathbf{a}_p, 0, 0, \dots)$, где $\mathbf{a}_p = (a_1, a_2, \dots, a_p)$ – p -вектор КЛП.

Для вычислений в рамках проведенного эксперимента было использовано авторское программное обеспечение Phoneme Training⁶ при установленной в нем частоте дискретизации $F = 8$ кГц голосового сигнала $x(t)$ и объеме $n = \tau F = 128$ выборок наблюдений ($\tau = 16$ мс). Источником аддитивной помехи $\eta(t)$ служил стандартный генератор (Excel) белого гауссова шума с математическим ожиданием 0 и дисперсией 1. Значение ОСШ q^2 регулировалось переменным усилением сигнала звука речи $s(t)$ на входе сумматора (1).

Как видно из рис. 1, во всех трех вариантах спектральной оценки (3) явно выражены по четыре форманты, из них доминируют по амплитуде первые две [16]. Их относительный уровень зависит от ОСШ. По мере уменьшения ОСШ форманты “размываются” вдоль оси частот. Частоты формант под действием фоновых шума также подвержены искажениям [14]. Однако для первых двух формант, как наиболее интенсивных составляющих спектра мощности голосового сигнала, частотные искажения минимальны [6, 12].

В подтверждение сказанного на рис. 2 представлены три соответствующие диаграммы КЛП. Они имеют схожую структуру при разной амплитуде флуктуаций. В варианте $q^2 = 20$ дБ амплитуда максимальна. Для сравнения здесь же показана диаграмма КЛП в отсутствие фонового шума

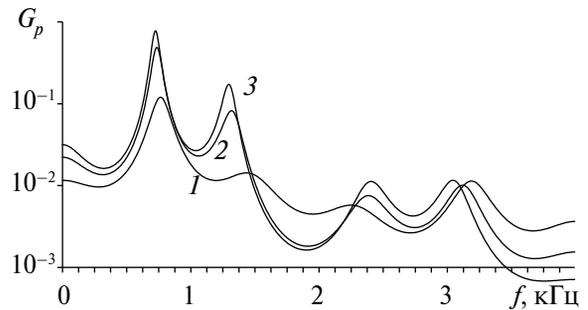


Рис. 1. Оценка огибающей СПМ (3) сигнала гласной фонемы “а” при ОСШ q^2 , равном 0 (1), 10 (2) и 20 дБ (3).

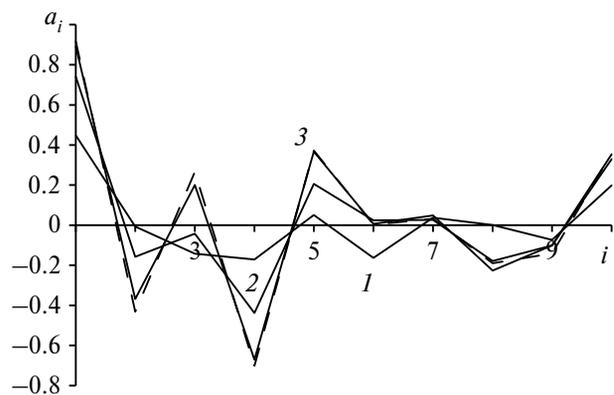


Рис. 2. Оценки КЛП сигнала фонемы “а” при ОСШ q^2 , равном 0 (1), 10 (2) и 20 дБ (3) в сопоставлении с вектором КЛП в отсутствие шума (пунктирная линия).

(пунктир). Она практически не отличается от диаграммы третьего варианта, отвечающего равенству $q^2 = 20$ дБ. По-видимому, именно этот уровень ОСШ можно считать пороговым применительно к конкретному звуку речи и конкретному диктору.

При $q^2 < 20$ дБ формируемый вектор КЛП (см. рис. 2) нуждается в корректировке. Она должна быть нацелена на избирательную по частоте регулировку усиления голосового сигнала в области его основных формант [20, 21]. Нетрудно понять, что в условиях априорной неопределенности это нетривиальная задача. Ее решение подсказывает рис. 3.

На рис. 3 представлено семейство графиков импульсной характеристики

$$h(t) = \sum_{i=1}^p a_i h(t-i) + \delta(t), \quad t = 0, 1, \dots, \quad (5)$$

формирующего фильтра (4), где символом $\delta(t)$ обозначен единичный импульс (δ -импульс) на входе. Графики построены по данным рис. 2 для трех разных значений ОСШ. Как видим, они сильно разнятся по длительности импульсных характеристик при том, что закон свободных колебаний в них слабо зависит от ОСШ. Отметим, что данное наблюдение точно отвечает свойствам аддитивного белого шума $\eta(t)$: его “формирующий фильтр”

⁵ Порядок $p = 10$ является стандартным для алгоритмов семейства CELP [5].

⁶ Программа Phoneme Training размещена в режиме открытого доступа на сайте <https://sites.google.com/site/frompldcreators/produkty-1/phonemetraining>.

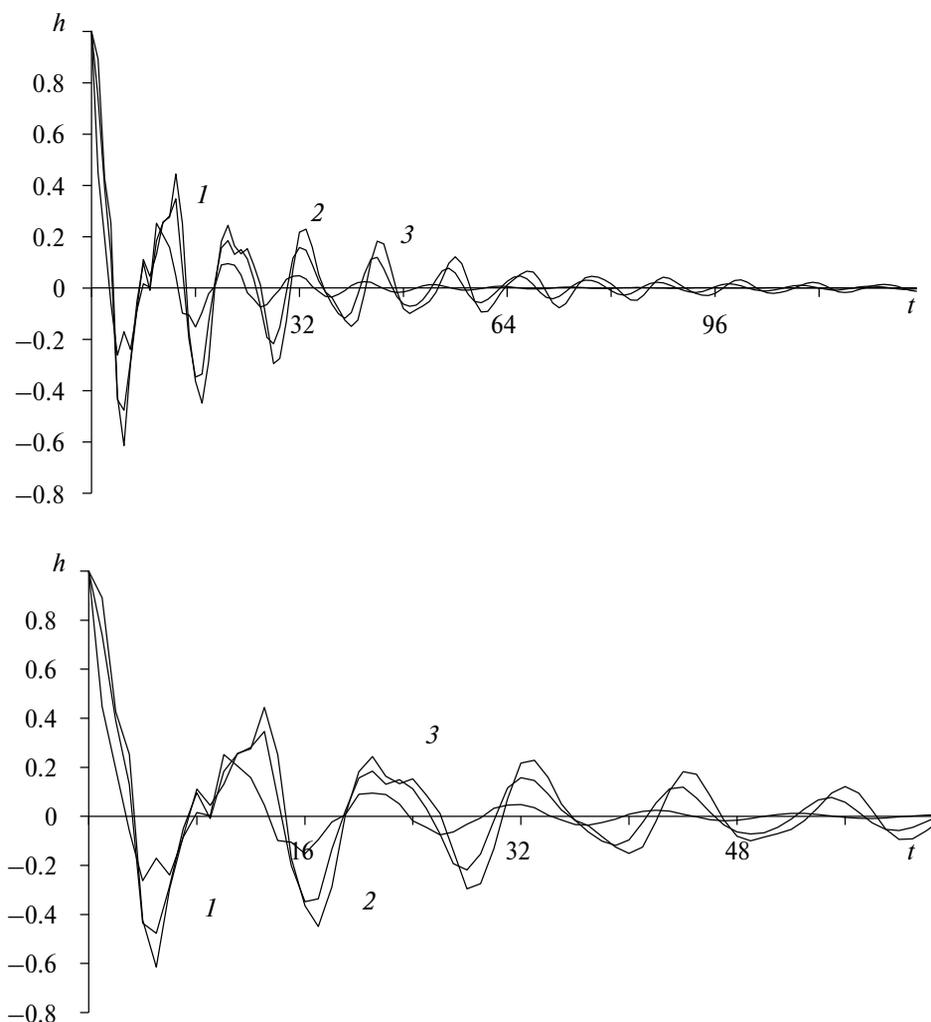


Рис. 3. Импульсная характеристика (5) формирующего фильтра (4) при ОСШ q^2 , равном 0 (1), 10 (2) и 20 дБ (3).

теоретически безынерционен и вырождается в “прямой провод” [11]. Вместе с тем следует учитывать [17, 19], что колебания импульсной характеристики формирующего фильтра — это отображение формант голосового сигнала во временной области. Чем более ярко (по амплитуде) выражены его форманты, тем медленнее затухают эти колебания. Отсюда можно заключить, что корректировка АР-модели (2) в общем случае может быть сведена к увеличению инерционности формирующего фильтра (4) путем корректировки его импульсной характеристики (5).

2. КОРРЕКТИРОВКА ИМПУЛЬСНОЙ ХАРАКТЕРИСТИКИ

Умножим импульсную характеристику (5) на монотонно возрастающую величину $g(t; c) \geq 1$:

$$h(t; c) = g(t; c)h(t), \quad t = 0, 1, \dots, \quad (6)$$

где c — регулируемый параметр, зависящий от ОСШ q^2 и целей корректировки. Например, это может быть экспоненциальная зависимость вида

$$g(t; c) = \exp(ct) \quad \text{при } c > 0$$

как бесконечно гладкая (дифференцируемая) функция своего аргумента t .

На рис. 4 представлены графики откорректированной импульсной характеристики (6) при равенстве ОСШ $q^2 = 0$ дБ для трех разных значений параметра c : 0.01, 0.03 и 0.05, а также в отсутствие корректировки при $c = 0$ (пунктир, который повторяет собой кривую 1 на рис. 3). Эффект в данном случае очевиден: во всех вариантах откорректированной импульсной характеристики (6) пропорционально параметру $c > 0$ увеличилась длительность свободных колебаний, а следовательно, увеличилась инерционность формирующего фильтра (4) в его откорректированном варианте.

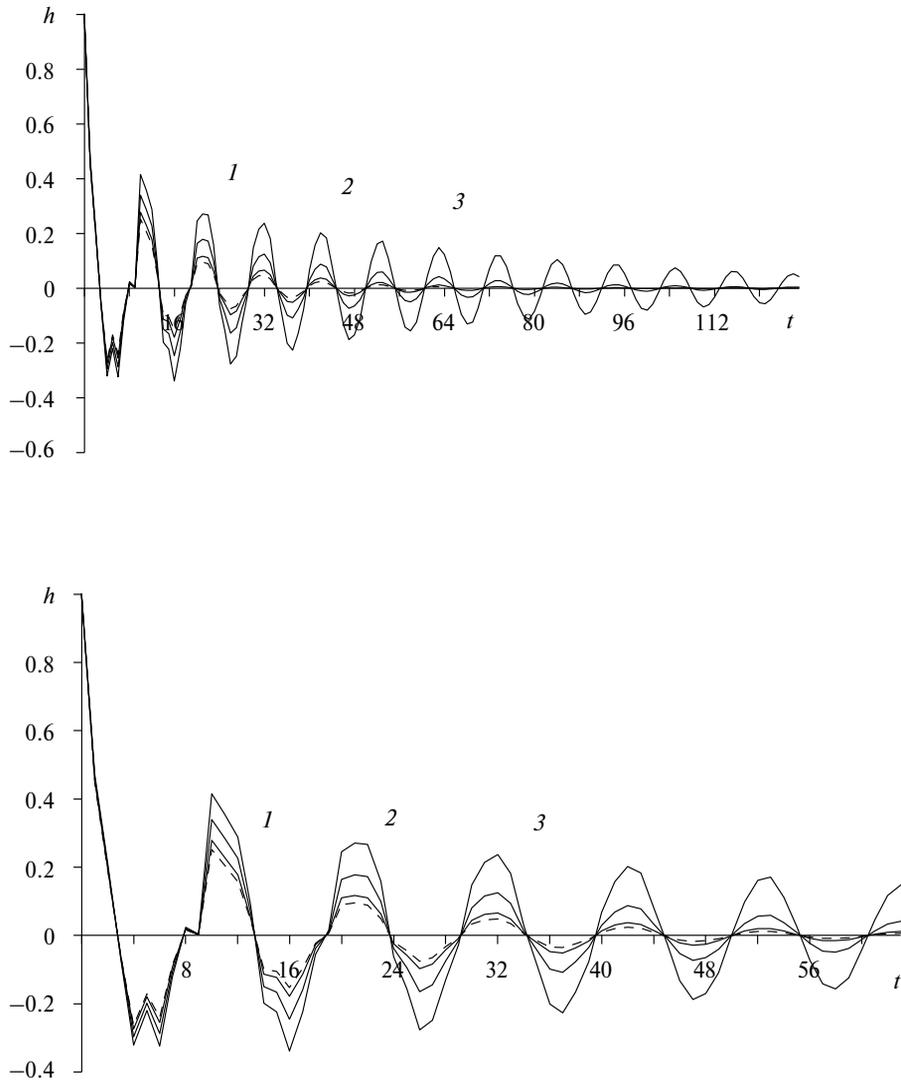


Рис. 4. Откорректированная импульсная характеристика (6) при $c = 0.01$ (1), 0.03 (2) и 0.05 (3) для случая равенства ОСШ $q^2 = 0$ дБ в сопоставлении с импульсной характеристикой (5) в отсутствие корректировки (пунктирная линия).

Указанный эффект подтверждает рис. 5, на котором графики огибающей СПМ откорректированного голосового сигнала построены по данным рис. 4, согласно выражению (3) и выражению комплексного коэффициента передачи откорректированного формирующего фильтра

$$K_p(jf; c) = \text{DFT}\{h(t; c)\}. \quad (7)$$

При этом для вычислений (7) использовался алгоритм быстрого преобразования Фурье (БПФ) размерностью $m = 2^{10} \gg n$. Для сравнения здесь же показана соответствующая оценка СПМ в отсутствие корректировки импульсной характеристики формирующего фильтра (4). Как видим, форманты в спектрах мощности откорректированных сигналов увеличились по амплитуде при сохранении практически неизменными своих частот. При этом первая форманта увеличилась более чем на 10 дБ. А как известно [16, 20], такой эффект служит

существенному улучшению разборчивости речи диктора.

Задача состоит теперь в пересчете откорректированной импульсной характеристики $h(t; c)$ в

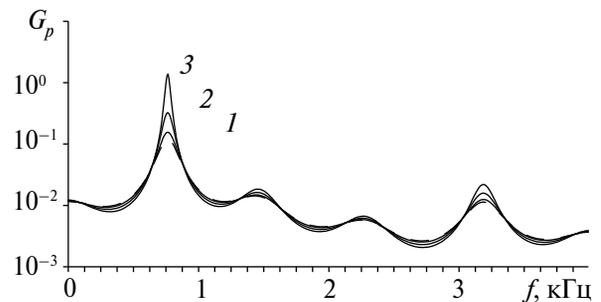


Рис. 5. Огибающая СПМ (3) синтезированного голосового сигнала при $c = 0.01$ (1), 0.03 (2) и 0.05 (3) для случая равенства ОСШ $q^2 = 0$ дБ и в отсутствие корректировки (пунктирная линия).

откорректированный вектор КЛП $\mathbf{a}_p(c)$ AP-модели голосового сигнала (2).

3. КОРРЕКТИРОВКА ВЕКТОРА КЛП

В соответствии с выражением (4) запишем равенство

$$b_i = \text{DFT}_i^{-1} \{TK_p^{-1}(jf)\} = \\ = T \int_{-0.5F}^{+0.5F} K_p^{-1}(jf) \exp\{+j2\pi ifT\} df, \quad i = 0, 1, \dots, p,$$

где $\text{DFT}^{-1}\{\cdot\}$ — оператор обратного ДФП. По аналогии с этим равенством для вектора коэффициентов откорректированного формирующего фильтра (8) будем иметь

$$\forall i \leq p: b_i(c) = \text{DFT}_i^{-1} \{TK_p^{-1}(jf; c)\}. \quad (8)$$

Соответствующий p -вектор $\mathbf{a}_p(c) = [a_1(c), a_2(c), \dots, a_p(c)]$ откорректированных КЛП определим согласно системе равенств

$$\forall i = \overline{1, p}: a_i(c) = -\overline{b_{i+1}(c)}. \quad (9)$$

Выражения (5)–(9) в совокупности определяют регулярный метод коррективки AP-модели (2) аддитивной смеси (1) голосового сигнала с фоновым шумом.

Для иллюстрации возможностей разработанного метода в табл. 1 представлены результаты его применения в условиях рассмотренного выше эксперимента (см. рис. 1–3) для случая равенства ОСШ $q^2 = 0$ дБ. Ее верхняя строка (соответствует равенству $c = 0$) составлена из первоначальных значений КЛП (см. рис. 3, кривая 1).

Отметим, что согласно рис. 4, во всех представленных в таблице вариантах коррективки КЛП обеспечен затухающий характер импульсной характеристики формирующего фильтра (4) и этим гарантирована устойчивость AP-модели (2) в смысле цифровой фильтрации [15, 22].

В подтверждение сказанного на рис. 6 представлено семейство сигналов фонемы “а”, синтезированных по данным табл. 1 согласно уравнению (2) с использованием в качестве сигнала возбуждения $\xi(t)$ последовательности δ -импульсов с частотой

повторения 100 Гц. При этом вне зависимости от значения параметра $c > 0$ синтезированный сигнал сохраняет свою тонкую структуру. Меняется лишь скорость затухания его внутривысокочастотных колебаний — в сторону большей или меньшей степени вокализации. Во многих случаях этот эффект имеет самостоятельное значение, например, с точки зрения мелодики речи диктора [2].

4. ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Результаты проведенного исследования имеют непосредственное отношение к проблеме помехоустойчивости системы ЦОР. Однако она далеко не исчерпывает собой множества острых проблем в области речевых коммуникаций [23, 24]. Примером могут служить проблемы устойчивости AP-модели и улучшения акустического качества синтезированного голосового сигнала [22, 25]. Амплитуда формант в этом случае может меняться в обоих направлениях: как в сторону увеличения, так и уменьшения — в зависимости от целей предпринимаемой коррективки. С указанной точки зрения представляют интерес два варианта вектора КЛП (9): исходного $\mathbf{a}_p(0) = (0.917124236; -0.430847898; 0.265832062; -0.701148961; 0.372491825; 0.002903572; 0.031471774; -0.189665721; -0.140846521; 0.356422549)$ и откорректированного $\mathbf{a}_p(-0.06) = (0.863715078; -0.382127806; 0.222041603; -0.551543307; 0.275948731; 0.002025753; 0.020678429; -0.117361998; -0.082078064; 0.195608842)$ в задаче синтеза сигнала гласной “а” в отсутствие фонового шума $\eta(t)$. Различия между ними относятся исключительно к тембру голоса диктора: во втором варианте он заметно приглушен. Это явно следует из сравнения двух графиков периодограммы Шустера [9]

$$G_N(f; c) = N^{-1} T |\text{DFT}\{y(t; c)\}|^2 \quad (10)$$

в качестве оценок СПМ двух синтезированных сигналов с параметрами $c = 0$ и $c = -0.06$ (рис. 7).

Обе оценки получены по последовательности из $N = 2^{10}$ отсчетов сигнала на выходе соответствующего (установленному значению параметра c) формирующего фильтра (7). При этом на вход фильтра в качестве сигнала возбуждения $\xi(t)$ подавалась последовательность δ -импульсов с частотой 100 Гц.

Таблица 1. Откорректированные значения КЛП $\{a_i(c)\}$

c	i									
	1	2	3	4	5	6	7	8	9	10
0	0.448039	-0.007	-0.14238	-0.17098	0.050631	-0.16325	0.037343	0.001193	-0.07275	0.198174
0.01	0.452542	-0.00714	-0.14671	-0.17796	0.053227	-0.17334	0.040051	0.001292	-0.0796	0.219016
0.03	0.461684	-0.00743	-0.15578	-0.19278	0.058825	-0.19544	0.046069	0.001516	-0.0953	0.267507
0.05	0.471011	-0.00773	-0.16542	-0.20883	0.065011	-0.22036	0.052992	0.001779	-0.1141	0.326733

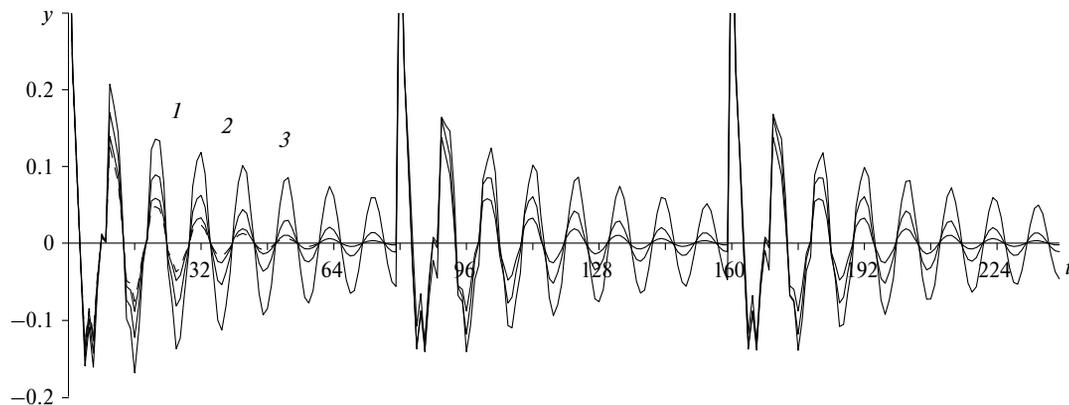


Рис. 6. Фрагменты синтезированного сигнала гласной фонемы “а” при $c = 0.01$ (1), 0.03 (2) и 0.05 (3) для случая равенства ОСШ $q^2 = 0$ дБ и в отсутствие коррективы (пунктирная линия).

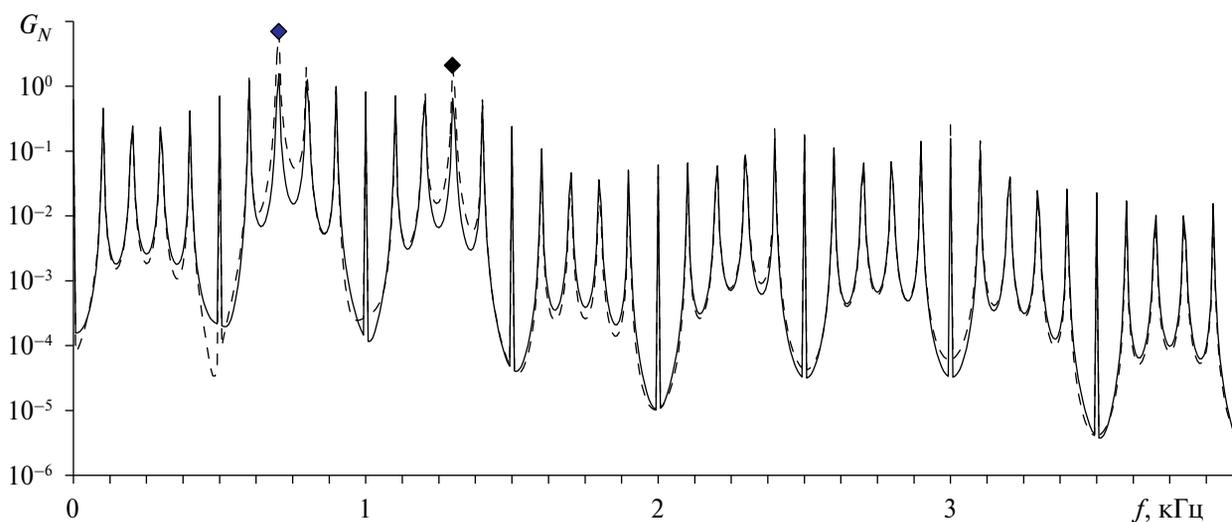


Рис. 7. Периодограмма Шустера (10) синтезированного согласно АР-модели (2) сигнала гласной фонемы “а” при $c = -0.06$ (сплошная кривая) и $c = 0$ (пунктирная линия).

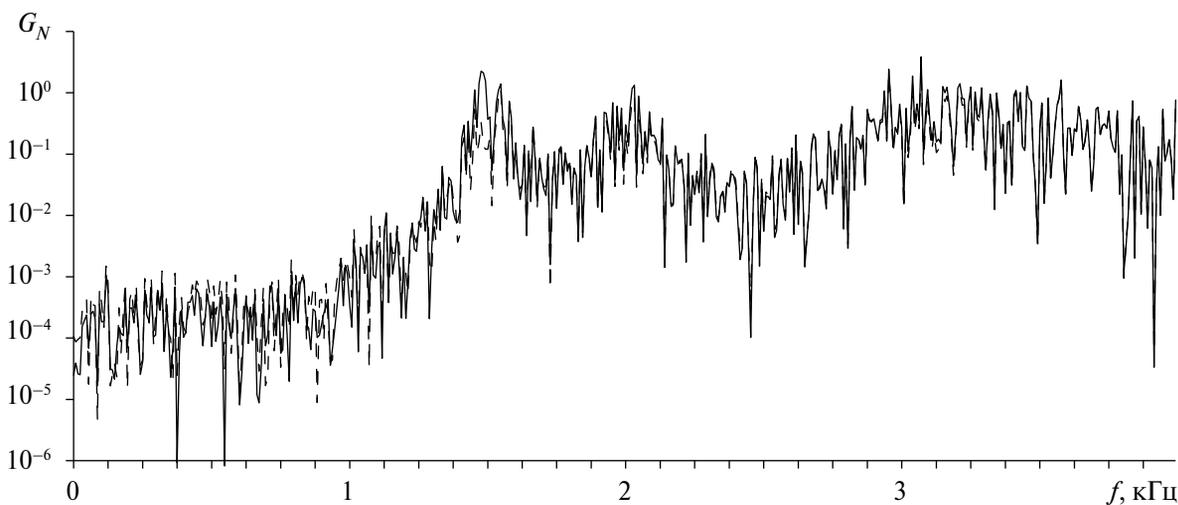


Рис. 8. Периодограмма Шустера (10) сигнала фрикативного звука речи “ш”, синтезированного согласно АР-модели (2) при $c = 0.06$ (сплошная кривая) и $c = 0$ (пунктирная линия).

Для вычислений ДФП использовался алгоритм БПФ размерностью $m = N = 2^{10}$. Как видим, в откорректированном варианте обе основные форманты ослаблены по амплитуде по сравнению с их исходными вариантами (отмечены знаком “ромб”). Это следствие знака “минус” перед значением параметра c в выражении (10).

Особо отметим, что область применения предложенного метода распространяется не только на гласные, но и на фрикативные звуки речи, где под формантами понимают не отдельные спектральные составляющие голосового сигнала, а целые области спектра мощности, в которых концентрируется большая часть его энергии [26]. Это подтверждает, в частности, проведенный автором статьи эксперимент с фонемой “ш” от контрольного диктора (рис. 8).

На рис. 8 представлены графики оценки СПМ (10) двух голосовых сигналов $y(t) = y(t, c)$ с параметрами $c = 0$ и 0.06 . Оба сигнала синтезированы согласно АР-модели (2) с использованием белого гауссова шума $\xi(t)$ в качестве возбуждения. Для вычислений периодограммы Шустера здесь, как и в предыдущем случае, применялся алгоритм БПФ размерностью $m = 2^{10}$. Как видим, в откорректированном варианте оценки спектра мощности амплитуда формант в данном случае ($c > 0$) возросла.

ЗАКЛЮЧЕНИЕ

Благодаря проведенному исследованию разработан регулярный метод корректировки АР-модели голосового сигнала для систем ЦОР со сжатием данных на основе линейного предсказания. В его основу положена оригинальная методика формантного анализа звука речи диктора через анализ колебаний импульсной характеристики формирующего фильтра.

К числу перспективных направлений исследований с применением разработанного метода относится область корректировки голосового сигнала при учете технических характеристик канала речевой связи, просодики речи диктора, а также селективных особенностей его голосового тракта.

Авторы данной работы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. *Rabiner L.R., Schafer R.W.* // Foundations and Trends in Signal Processing. 2007. V. 1. № 1–2. P. 1. <https://doi.org/10.1561/2000000001>
2. *O’Shaughnessy D.* // J. Audio. Speech. Music Processing. 2023. V. 8. <https://doi.org/10.1186/s13636-023-00274-x>
3. *Savchenko V.V.* // Radioelectron. Commun. Systems. 2021. V. 64. № 11. P. 592. <https://doi.org/10.3103/S0735272721110030>
4. *Gibson J.* // Information. 2019. V. 10. № 5. 179. <https://doi.org/10.3390/info10050179>
5. *Chaouch H., Merazka F., Marthon Ph.* // Speech Commun. 2019. V. 108. P. 33. <https://doi.org/10.1016/j.specom.2019.02.002>
6. *Савченко В.В., Савченко Л.В.* // Измерит. техника. 2019. № 9. С. 59. <https://doi.org/10.32446/0368-1025it.2019-9-59-64>
7. *Candan Ç.* // Signal Processing. 2020. V. 166. № 10. P. 107256. <https://doi.org/10.1016/j.sigpro.2019.107256>
8. *Semenov V.Yu.* // J. Automation and Inform. Sci. 2019. V. 51. № 2. P. 30. <https://doi.org/10.1615/JAutomatInfScien.v51.i2.40>
9. *Marple S.L.* Digital Spectral Analysis with Applications. 2-nd ed. Mineola: Dover Publ., 2019.
10. *Burg J.P.* Maximum entropy spectral analysis. PhD Thesis. Stanford Univ., 1975.
11. *Magi C., Pohjalainen J., Bäckström T., Alku P.* // Speech Commun. 2009. V. 51. № 5. P. 401. <https://doi.org/10.1016/j.specom.2008.12.005>
12. *Rout J.K., Pradhan G.* // Speech Commun. 2022. V. 144. P. 101. <https://doi.org/10.1016/j.specom.2022.09.004>
13. *Deng F., Bao Ch.* // Speech Commun. 2016. V. 79. P. 30. <https://doi.org/10.1016/j.specom.2016.02.006>
14. *Савченко В.В., Савченко А. В.* // Измерит. техника. 2020. № 11. С. 65. <https://doi.org/10.32446/0368-1025it.2020-11-65-72>
15. *Савченко В.В.* // ПЭ. 2023. Т. 68. № 2. С. 138. <https://doi.org/10.31857/S0033849423020122>
16. *Kathiresan Th., Maurer D., Suter H., Dellwo V.* // J. Acoust. Soc. Amer. 2018. V. 143. № 3. P. 1919. <https://doi.org/10.1121/1.5036258>
17. *Ngo Th., Kubo R., Akagi M.* // Speech Commun. 2021. V. 135. P. 11. <https://doi.org/10.1016/j.specom.2021.09.004>
18. *Palaparthi A., Titze I. R.* // Speech Commun. 2020. V. 123. P. 98. <https://doi.org/10.1016/j.specom.2020.07.003>
19. *Sadasivan J., Seelamantula Ch.S., Muraka N.R.* // Speech Commun. 2020. V. 116. P. 12. <https://doi.org/10.1016/j.specom.2019.11.001>
20. *Gustafsson Ph.U., Laukka P., Lindholm T.* // Speech Commun. 2023. V. 146. P. 82. <https://doi.org/10.1016/j.specom.2022.12.001>
21. *Ito M., Ohara K., Ito A., Yano M.* // Proc. Interspeech. 2010. V. 2490. <https://doi.org/10.21437/Interspeech.2010-669>
22. *Arun-Sankar M.S., Sathidevi P. S.* // Heliyon. 2019. V. 5. № 5. P. e01820. <https://doi.org/10.1016/j.heliyon.2019.e01820>
23. *Narendra N.P., Alku P.* // Speech Commun. 2019. V. 110. P. 47. <https://doi.org/10.1016/j.specom.2019.04.003>

24. Alku P., Kadiri S.R., Gowda D. // *Computer Speech & Language*. 2023. V. 81. № 10. P. 101515.
<https://doi.org/10.1016/j.csl.2023.101515>
25. Sadok S., Leglaive S., Girin L. et al. // *Speech Commun.* 2023. V. 148. P. 53.
<https://doi.org/10.1016/j.specom.2023.02.005>
26. Nguyen D.D., Chacon A., Payten Ch.L. et al. // *Int. J. Language & Commun. Disorders*. 2022. V. 57. № 2. P. 366.
<https://doi.org/10.1111/1460-6984.12705>

LINEAR PREDICTION COEFFICIENTS CORRECTION METHOD FOR DIGITAL SPEECH PROCESSING SYSTEMS WITH DATA COMPRESSION BASED ON THE AUTOREGRESSIVE MODEL OF A VOICE SIGNAL

V. V. Savchenko^{a, *}, L. V. Savchenko^b

^aEditorial office of the journal "Radio Engineering and Electronics", Mokhovaya St., 11, bldg. 7, Moscow, 125009 Russian Federation

^bNational Research University Higher School of Economics, B. Pecherskaya St., 25, Nizhny Novgorod, 603155 Russian Federation

*E-mail: vvsavchenko@yandex.ru

Received September 18, 2023; revised December 18, 2023; accepted January 16, 2024

The problem of distortion of the autoregressive model of the voice signal under the influence of additive background noise in digital speech processing systems with data compression based on linear prediction is considered. In the frequency domain, these distortions are observed in the weakening of the main formants responsible for the intelligibility of the speaker's speech. To compensate for formant attenuation, it is proposed to modify the parameters of the autoregressive model (linear prediction coefficients) using the impulse response of a recursive shaping filter. Along with the amplitude amplification of the formants, their frequencies remain unchanged to make the speaker's voice recognizable. The effectiveness of the method was studied experimentally using specially developed software. Based on the experimental results, conclusions were drawn about a significant increase in the relative level of formants in the power spectrum of the corrected voice signal.

Keywords: signal theory, voice signal, digital speech processing, digital speech transmission, spectral analysis, power spectral density, discrete spectral modeling, autoregressive model, all-pole model