Научная статья
УДК 004.056
DOI: 10.26583/bit.2025.2.03

ФИЗИОЛОГИЧЕСКИЕ АСПЕКТЫ ПОСТРОЕНИЯ СОНОГРАММ И РЕКОНСТРУКЦИИ СПЕКТРА ИСКАЖЕННЫХ РЕЧЕВЫХ ВОКАЛИЗМОВ

Михаил В. Алюшин[1], Александр М. Алюшин[2], Сергей В. Дворянкин[3], Никита С. Дворянкин[4]
[1–4]Национальный исследовательский ядерный университет «МИФИ», Каширское ш., 31, Москва, 115409, Россия
[2,3]Московский государственный лингвистический университет, ул. Остоженка, 38, стр. 1, Москва, 119034, Россия
[1]MVAlyushin@mephi.ru, https://orcid.org/0000-0001-7806-3739
[2]AMAlyushin@mephi.ru, https://orcid.org/0000-0003-1722-0598
[3]SVDvoryankin@mephi.ru, https://orcid.org/0000-0001-6908-0676
[4]nik.dvrn@gmail.com, https://orcid.org/0000-0002-1580-7179

Аннотация. Целью исследования является обоснование физиологического подхода к спектральному анализу-синтезу речевого сигнала (РС), предполагающему учет уникальных для каждого человека биометрических параметров, обусловленных физической природой его речеобразующего тракта, необходимых для решения задач защиты речевой информации (РИ) от фальсификации и подделок, а также для восстановления речевой разборчивости (РР) на искаженных участках голосовых интерфейсов. Выделены основные группы используемых для этого физических параметров артикуляционной системы человека. Проанализировано влияние непостоянной во времени формы, огибающей гармонических составляющих РС на получаемое частотное разрешение при выполнении кратковременного преобразования Фурье (КПФ, STFT) с взвешивающим окном Гаусса. Представлено сравнение разрешающей способности преобразования Фурье с окном Гаусса при спектральном анализе-синтезе синусоидальных сигналов с различными огибающими функциями. Показана возможность применения полученных результатов для реконструкции гармонической структуры спектра искаженных участков речевых вокализмов (РВ).

Ключевые слова: *спектральный анализ-синтез, защита речевой информации, реконструкция речевого сигнала, восстановление разборчивости речи, речевые вокализмы.*

Для цитирования: *Алюшин, Михаил В. и др. Физиологические аспекты построения сонограмм и реконструкции спектра искаженных речевых вокализмов. Безопасность информационных технологий, [S.l.], т. 32, № 2, с. 32–47, 2025. ISSN 2074-7136. URL: https://bit.spels.ru/index.php/bit/article/view/1780. DOI: 10.26583/bit.2025.2.03.*

Scientific article

PHYSIOLOGICAL ASPECTS FOR SONOGRAMS BUILDING AND SPECTRUM RESTORE DISTORTED SPEECH VOCALIZATIONS

Mikhail V. Alyushin[1], Alexander M. Alyushin[2], Sergey V. Dvoryankin[3], Nikita S. Dvoryankin[4]
[1–4]National Nuclear Research University MEPHI (Moscow Engineering Physics Institute), Kashirskoe sh., 31, Moscow, 115409, Russia
[2,3]Moscow State Linguistic University, Ostojenka str., 38/1, Moscow, 119034, Russia
[1]MVAlyushin@mephi.ru, https://orcid.org/0000-0001-7806-3739
[2]AMAlyushin@mephi.ru, https://orcid.org/0000-0003-1722-0598

[3]*SVDvoryankin@mephi.ru, https://orcid.org/0000-0001-6908-0676*
[4]*nik.dvrn@gmail.com, https://orcid.org/0000-0002-1580-7179*

Abstract. The aim of the study is to substantiate the physiological approach to spectral analysis-synthesis of the speech signal (RS), which involves taking into account the biometric parameters unique for each person due to the physical nature of his speech-forming tract, necessary for solving the problems of protection of speech information (SI) from falsification and forgery, as well as for restoration of speech intelligibility (RS) at distorted areas of voice interfaces. The main groups of physical parameters of the human articulatory system used for this purpose are highlighted. The influence of the time-varying shape of the envelope of the harmonic components of the RS on the obtained frequency resolution when performing a short-time Fourier transform (STFT or CPF) with a Gaussian weighting window is analyzed. A comparison of the resolution of the Fourier transform with a Gaussian window for spectral analysis-synthesis of sinusoidal signals with different envelope functions is presented. The possibility of application of the obtained results for reconstruction of the harmonic structure of distorted parts of speech vocalizations (SVs) is shown.

Keywords: *spectral analysis-synthesis, speech information protection, speech signal reconstruction, speech intelligibility restoration, speech vocalizations.*

For citation: *Alyushin, Mikhail V. et al. Physiological aspects for sonograms building and spectrum restore distorted speech vocalizations. IT Security (Russia), [S.l.], v. 32, no. 2, p. 32–47, 2025. ISSN 2074-7136. URL: https://bit.spels.ru/index.php/bit/article/view/1780. DOI: 10.26583/bit.2025.2.03.*

Введение

Последние достижения в области компьютерной шумоочистки, синтеза речевых сообщений, имитирующих голос практически любого человека с помощью искусственных нейросетей и др. [1, 2], обусловливают актуальность задачи использования дополнительных, индивидуальных биометрических характеристик речевого сигнала (РС) в управляющих, информационных и телекоммуникационных системах для восстановления речевой разборчивости (РР), своевременного выявления случаев голосовых подделок и фальсификаций, подтверждения подлинности речевых команд.

Показательным примером в этом плане может служить современная технология защиты важных финансовых и юридических документов с помощью так называемой речевой подписи (РП) [3–5].

Индивидуальные характеристики РС определяются уникальным строением речеобразующего тракта каждого человека. Так, с физической точки зрения РС представляет собой результат воздействия импульсов воздушного потока, формируемого голосовыми связками человека, на его артикуляционную систему (рис. 1) [6–8], которая обуславливает индивидуальную окраску голоса. По этой причине именно проявление в голосе уникальных физических характеристик артикуляционной системы целесообразно использовать в качестве биометрически информативных параметров.

В табл. 1 представлен перечень основных физических характеристик артикуляционной системы человека, проявляющихся в индивидуальных особенностях его речи, а также показана специфика их использования в настоящее время и даны оценки возможности и целесообразности их применения для повышения биометрической информативности РС.

Артикуляционная система человека обладает целым набором уникальных персонализированных характеристик Ф1-Ф3, обусловливающих высокую потенциальную биометрическую информационность формируемого ею РС. Ряд биометрических характеристик традиционно используется при фоноскопическом анализе РС [9].

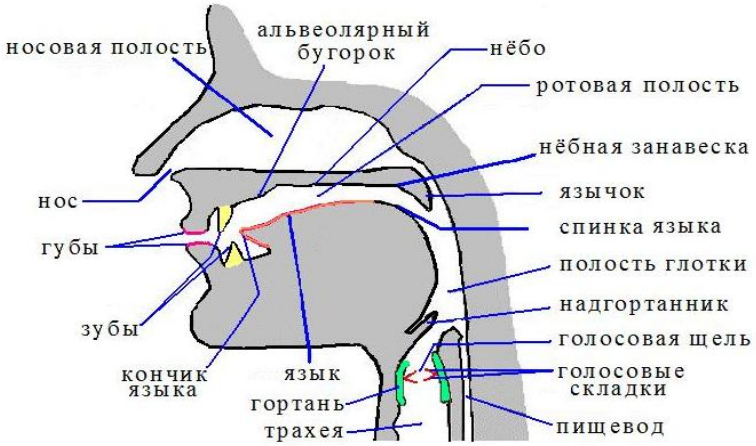


Рис. 1. Строение артикуляционной системы человека [7]

Таблица 1. Физические характеристики артикуляционной системы человека

№	Физические параметры артикуляционной системы человека (группа физических параметров)	Влияние на уникальные характеристики РС	Специфика использования в настоящее время	Возможность применения для защиты РС от подделок и восстановления РР
1	Форма и объем акустических резонансных полостей артикуляционной системы человека (Ф1)	Обусловливают спектр резонансных частот, в том числе, частоту основного тона (ЧОТ).	Используются при моделировании и спектральных преобразованиях, синтезе РП [3–8].	Широко используются, в том числе, для создания сонограммы РП, реконструкции искаженных участков синтезируемого РС
2	Физическое состояние стенок артикуляционной системы (толщина, упругость, звуковая «прозрачность» стенок акустических резонансных полостей) (Ф2)	Обусловливает скорость спада амплитуды акустических колебаний после их возбуждения.	Используется при 1D, 2D и 3D моделировании процесса речеобразования, реконструкции искаженных РС, обнаружения признаков натуральности [6–8].	Использование для выявления подделок РС, восстановления РР, в том числе для сонограммы РП, возможно и целесообразно.
3	Параметры акустической связи с источником возбуждения (голосовыми связками) (Ф3)	Обусловливают временную задержку и начальную фазу колебаний, возникающих в различных разделах речевого тракта .	Используется при 1D, 2D и 3D моделировании процесса речеобразования, реконструкции искаженных РС, обнаружения признаков естественности [6–8].	Использование для выявления подделок РС, восстановления РР, в том числе, для сонограммы РП, возможно и целесообразно.

Биометрическая информативная ограниченность применяемых в настоящее время технологий обработки, синтеза и анализа РС снижает его защищенность от подделок и фальсификаций, а также возможность восстановления РР искаженных шумами и помехами участков РС. Это прежде всего относится к группам биометрических параметров Ф2 и Ф3.

Так, например, применяемое для создания РП спектральное преобразование кратковременного Фурье-анализа (КФА, STFT) для некоторых моделей РС не позволяет в полной мере передать индивидуальные биометрические характеристики, обусловленные физическими факторами Ф2 и Ф3. Этот факт ограничивает возможность реконструкции гармонической структуры по оставшимся на сонограмме следам фонообъектов, что в итоге может привести к синтезу неестественно и неправдоподобно звучащего РС.

Целью настоящего исследования является обоснование физического подхода к спектральному анализу-синтезу РС, используемому в системах защиты РИ и предполагающему учет уникальных для каждого человека биометрических параметров, обусловленных физической природой его речеобразующего тракта.

Для оценки степени влияния реальной формы РС, отображающей воздействие физических факторов Ф2 и Ф3 на информативность получаемой спектрограммы, а также возможность дальнейшей её обработки и последующего синтеза по ней речеподобного сигнала (РПС) с заданными свойствами, проанализируем процедуру спектрального анализа, реализуемого с помощью КФА для различных моделей РС.

1. Эффект расширения спектральных линий при использовании преобразования Фурье для простого тонального сигнала

При преобразовании волны акустического сигнала в графический образ спектрограммы возникает искажение изображения − оно размывается по вертикали. Это можно объяснить следующим образом. При формировании сонограммы из амплитудно-временного представления звукового сигнала в каждую часть спектральной области преобразуется некоторый фрагмент исходного сигнала, взятый на небольшом участке времени. Можно полагать, что он состоит из совокупности микрогармоник, некоторые из которых имеют достаточно близкие частоты. В процессе построения динамической спектрограммы частоты этих близких микрогармоник будут располагаться по соседству, а их спектры могут перекрывать друг друга (рис. 2) и иногда даже сливаться.

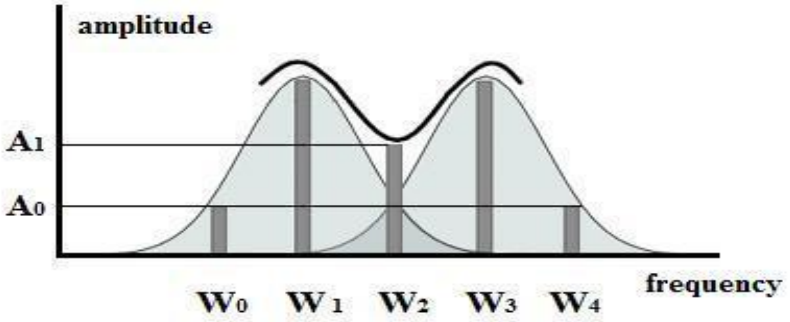
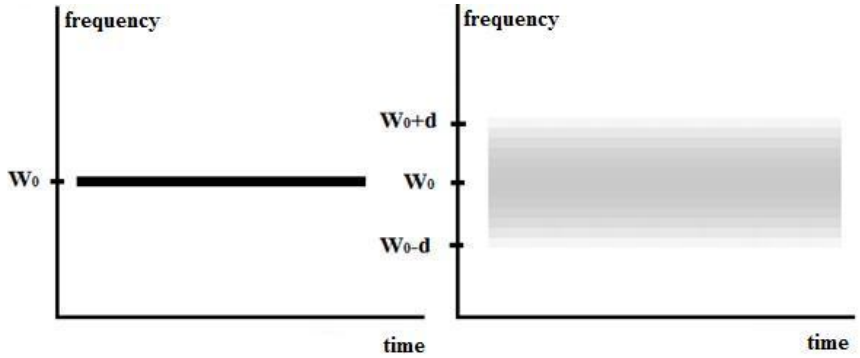


Рис. 2. Наложение спектров микрогармоник близких частот при некорректно выбранном разрешении

В результате истинные спектральные значения близких частот будут искажены.

На динамической сонограмме это отразится в виде размазывания узкой спектральной линии на частотно-временной плоскости в некоторую область – полосу с

шириной, состоящей из нескольких пикселей в зависимости от величины шага частотной сетки (рис. 3). Трек локальных максимумов (ЛМ) посередине этой полосы будет хорошо соответствовать исходной спектральной линии микрогармоники с частотой W0. Этот параметр частоты вместе с амплитудой и синтетической фазой сможет участвовать в формировании гармонической и формантной составляющих искаженного шумами и помехами речевых вокализмов (РВ) [1]. Однако из-за эффекта расширения спектральных линий на это потребуется значительное количество временных и вычислительных ресурсов, чтобы рассчитать эти параметры с точностью, необходимой для восстановления РР или создания речевого клона.



a) *b)*

Рис. 3. Частотная составляющая сигнала (a) и ее сонограмма (b)

Для поиска путей нейтрализации эффекта расширения спектральных линий оценим получаемую ширину единственной спектральной линии в сонограмме, построенной в результате применения преобразования Фурье для простого синусоидального сигнала

$$S(t) = A \cdot \cos(\omega_0 t + \varphi_0). \quad (1)$$

Рассмотрим преобразование Фурье

$$S(\omega) = \int_{-\infty}^{+\infty} S(t) \cdot G(t) \cdot e^{-i\omega t} dt \quad (2)$$

с нормированным окном Гаусса $G(t)$

$$G(t) = \frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{t^2}{2\delta^2{}_t}}. \quad (3)$$

Получаем:

$$S(\omega) = \int_{-\infty}^{+\infty} A \cdot \cos(\omega_0 t + \phi_0) \cdot (\frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{t^2}{2\delta^2{}_t}}) \cdot e^{-i\omega t} dt. \quad (4)$$

Заменяем функцию косинуса на полусумму экспонент в степени i, получаем:

$$S(\omega) = \frac{A}{\sqrt{2\pi}\delta^2{}_t} \int_{-\infty}^{+\infty} \frac{e^{i(\omega_0 t + \varphi_0)} + e^{-i(\omega_0 t + \varphi_0)}}{2} \cdot e^{-\frac{t^2}{2\delta^2{}_t} - i\omega t} dt. \quad (5)$$

Используя значение известного табличного интеграла на интервале $-\infty<t<+\infty$

$$\int_{-\infty}^{+\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}, \tag{6}$$

в итоге получаем выражение для спектра анализируемого сигнала:

$$S(\omega) = \frac{A}{2} \cdot e^{i\varphi_0 - \frac{(\omega-\omega_0)^2}{2\delta^2{}_\omega}} + \frac{A}{2} \cdot e^{-i\varphi_0 - \frac{(\omega+\omega_0)^2}{2\delta^2{}_\omega}}, \tag{7}$$

где

$$\delta_t = 1/\delta_\omega. \tag{8}$$

Спектр сигнала (1), полученный с помощью преобразования Фурье (2) с оконной функцией Гаусса (3), содержит две значимые зеркальные области с центрами соответственно в точках ω_0 и $-\omega_0$:

$$S(\omega) = S^-(\omega) + S^+(\omega), \tag{9}$$

где

$$S^-(\omega) = \frac{A}{2} \cdot e^{i\varphi_0 - \frac{(\omega-\omega_0)^2}{2\delta^2{}_\omega}}, \; S^+(\omega) = \frac{A}{2} \cdot e^{-i\phi_0 - \frac{(\omega+\omega_0)^2}{2\delta_\omega^2}}. \tag{10}$$

Функции $S^-(\omega)$ и $S^+(\omega)$ (10) имеют одинаковую амплитуду:

$$\left|S^-(\omega)\right| = \left|S^+(\omega)\right| = \frac{A}{2} \cdot e^{-\frac{(\omega-\omega_0)^2}{2\delta_\omega^2}}. \tag{11}$$

Обычно на практике рассматривают только одну компоненту спектра, не рассматривая зеркальную частоту $-\omega_0$:

$$S(\omega) = S^-(\omega). \tag{12}$$

При этом фаза косинуса (1) сохраняется:

$$arctg\left(\frac{\mathrm{Im}(S^-(\omega))}{\mathrm{Re}(S^-(\omega))}\right) = \phi_0. \tag{13}$$

Полученное выражение (11) позволяет определить ширину получаемой спектральной линии:

$$\Delta_\omega = 2\delta_\omega. \tag{14}$$

Таким образом, спектральное преобразование Фурье тонального сигнала с оконной функцией Гаусса шириной во временной области $\Delta_t = 2\delta_t$ приводит к расширению спектральной линии в частотной области до величины Δ_ω (14).

В дальнейшем будем рассматривать такую оценку в качестве базовой при сравнении результатов спектрального преобразования временных сигналов *S(t)* более сложной формы.

2. Оценка расширения спектральной линии временного сигнала со спадающей амплитудой

Для выявления эффекта дополнительного расширения спектральной линии при анализе синусоидальных составляющих РС с непостоянной амплитудой рассмотрим результаты спектрального преобразования (2) для синусоидального сигнала со спадающей амплитудой:

$$S(t) = A \cdot e^{-\alpha^2 x^2} \cdot \cos(\omega_0 t + \phi_0). \tag{15}$$

Для упрощения аналитических преобразований рассматриваем симметричную относительно точки t=0 экспоненциальную функцию спада.

Проведя аналогичные (2)–(7) преобразования, получаем для составляющих $S_e^-(\omega)$ и $S_e^+(\omega)$ (9):

$$S_e^-(\omega) = \frac{A}{2} \cdot \frac{1}{\sqrt{(1+2\alpha^2\delta_t^2)}} e^{i\phi_0 - \frac{(\omega-\omega_0)^2}{2\delta_\omega^2(1+2\alpha^2\delta_t^2)}}, \quad (16)$$

$$S_e^+(\omega) = \frac{A}{2} \cdot \frac{1}{\sqrt{(1+2\alpha^2\delta_t^2)}} e^{-i\phi_0 - \frac{(\omega+\omega_0)^2}{2\delta_\omega^2(1+2\alpha^2\delta_t^2)}}. \quad (17)$$

Амплитуды полученных компонентов спектра также имеют одинаковую величину (9):

$$\left|S_e^-(\omega)\right| = \left|S_e^+(\omega)\right| = \frac{A}{2} \cdot \frac{1}{\sqrt{(1+2\alpha^2\delta_t^2)}} e^{-\frac{(\omega-\omega_0)^2}{2\delta_\omega^2(1+2\alpha^2\delta_t^2)}}. \quad (18)$$

В этом случае также происходит сохранение фазы исходного колебания:

$$arctg\left(\frac{\mathrm{Im}(S_e^-(\omega))}{\mathrm{Re}(S_e^-(\omega))}\right) = \phi_0. \quad (19)$$

Из полученного выражения (18) видно, что происходит «размазывание» спектральной линии. При этом по сравнению с базовой шириной спектральной линии (14) амплитуда спектральной линии сложного входного сигнала стала зависеть от параметров α и δ_t и уменьшилась на величину (10), (18):

$$\Delta A = \frac{A}{2}\left(\frac{1}{\sqrt{(1+2\alpha^2\delta_t^2)}} - 1\right). \quad (20)$$

При α=0 (спад амплитуды входного сигнала отсутствует) выражение для амплитуды спектральной линии (20) преобразуется в базовое значение $A/2$, определяемое уравнениями (9), (10). Соответственно в этом случае ΔA=0.

Одновременно с уменьшением амплитуды спектральной линии происходит ее расширение на величину:

$$\Delta_\omega^{'} = 2\delta_\omega(1+2\alpha^2\delta_t^2) - 2\delta_\omega = 2\delta_\omega \cdot 2\alpha^2\delta_t^2. \quad (21)$$

С учетом (8) получаем:

$$\Delta_\omega^{'} = 4\frac{\alpha^2}{\delta_\omega} = 4\alpha^2\delta_t. \quad (22)$$

На рис. 4 показаны эффекты уменьшения амплитуды спектральной линии и ее расширения.

Таким образом, значение величины α^2, определяющей скорость спада амплитуды синусоидального колебания (задается декрементом затухания) имеет принципиальное значение при выполнении спектрального преобразования, так как определяет увеличение ширины спектральной линии и уменьшение ее амплитуды.

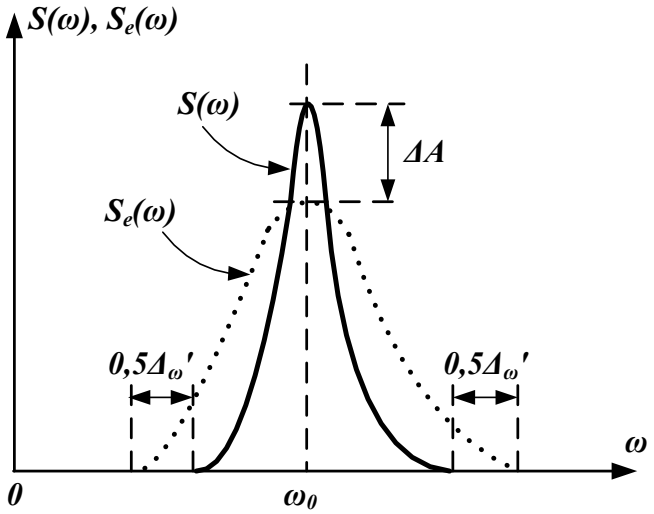


Рис. 4. Сравнение спектров

3. Физический смысл влияния спада амплитуды сигнала на параметры спектральной линии

На рис. 5 представлена иллюстрация эффекта влияния спада амплитуды одночастотного сигнала на параметры получаемой спектральной линии.

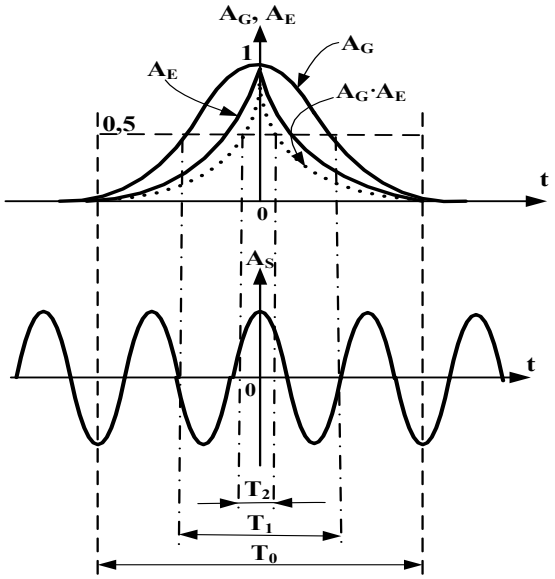


Рис. 5. Физическая интерпретация эффекта изменения характеристик спектра

Физический смысл влияния спада амплитуды на характеристики, получаемые в результате преобразования Фурье с окном Гаусса спектральной линии, заключается в сужении ширины эффективного временного окна анализа. Так, если при обычном преобразовании Фурье временное окно анализа имеет величину T_0, то при использовании окна Гаусса AG размер эффективного окна, определяющего 75% энергии гармоник (оценивается на уровне 0,5 амплитуды функции Гаусса) составляет величину $T_1<T_0$.

Значительно хуже дело обстоит при наличии спада амплитуды сигнала по экспоненциальной функции AE (15). Величина эффективного временного окна в этом случае составляет всего $T_2<T_1<T_0$. Принимая во внимание соотношение (8), получаем

оценку для эквивалентного значения ширины временного окна спектрального анализа входного сигнала со спадающей амплитудой (20):

$$\Delta_t^e = \frac{\delta_t}{2(1+2\alpha^2\delta_t^2)} \cdot \tag{23}$$

Спектральный анализ временного сигнала (15) с окном Гаусса может быть представлен как спектральный анализ Фурье синусоидального сигнала с постоянной амплитудой с использованием нового окна, обладающего другими свойствами:

$$G^e(t) = e^{-\alpha^2 x^2} \cdot \frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{t^2}{2\delta_t^2}} . \tag{24}$$

Таким образом, использование классического окна Гаусса для анализа сигнала (15) может привести к ухудшению разрешающей способности получаемого спектра.

4. Особенности спектрального анализа гармонизированного фрагмента речи

На рис. 6 показан гармонизированный фрагмент РС, речевой вокализм,

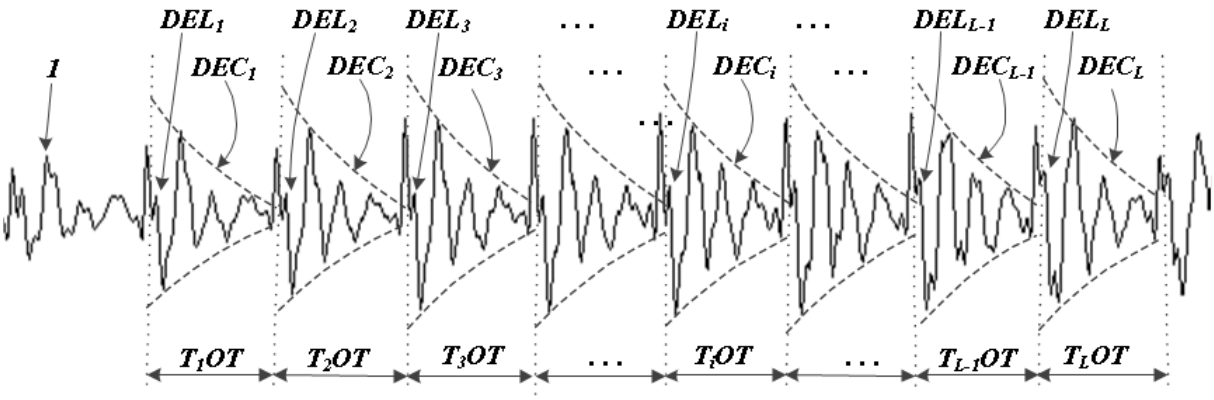


Рис. 6. Фрагмент вокализованного участка РС [1]:
где ***T_1OT-T_LOT*** *– временные периоды ОТ;* ***DEC_1-DEC_L*** *– интегральный спад амплитуды РС;* ***DEL_1-DEL_L*** *– момент времени формирования одной из высокочастотных гармоник.*

Рассмотрим спектральное преобразование Фурье, ширина временного окна Гаусса которого совпадает с периодом основного тона (РС). Будем использовать следующую модель фрагмента РС:

$$S(t) = A \cdot e^{-\beta x} \cdot \cos(\omega_0 t + \varphi_0) . \tag{25}$$

На рис. 7 показано относительное положение на временной оси окна спектрального анализа и функции Гаусса для рассматриваемого случая.

Рассмотрим преобразование Фурье фрагмента РС (25):

$$S(\omega) = \int_0^{+\infty} A \cdot e^{-\beta x} \cdot \cos(\omega_0 t + \varphi_0) \cdot \left(\frac{1}{\sqrt{2\pi}\delta_t} e^{-\frac{(t-0,5T_0)^2}{2\delta_t^2}} \right) \cdot e^{-i\omega t} dt . \tag{26}$$

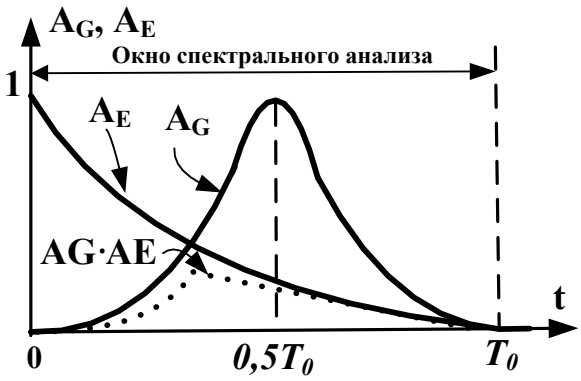


Рис. 7. Взаимное расположение окна анализа и функции Гаусса

Проведя аналитические преобразования, аналогичные (2)-(7) и учитывая, что с большой степенью точности

$$\int_0^{+\infty} e^{-a(x-0,5T)^2} d(x-0,5T_0) \cong \sqrt{\frac{\pi}{a}}, \tag{27}$$

получаем для компонент спектра:

$$S_e^-(\omega) = \frac{A}{2} \cdot e^{i(\phi_0-(\omega-\omega_0)\left(\delta_t^2-\frac{T_0}{2}\right)} \cdot e^{-\frac{(\omega-\omega_0)^2}{2\delta_\omega^2}} \cdot e^{-\beta\left(\frac{T_0}{2}-\frac{\delta_t^2}{2}\beta\right)},$$

$$S_e^+(\omega) = \frac{A}{2} \cdot e^{-i(\phi_0-(\omega-\omega_0)\left(\delta_t^2-\frac{T_0}{2}\right)} \cdot e^{-\frac{(\omega+\omega_0)^2}{2\delta_\omega^2}} \cdot e^{-\beta\left(\frac{T_0}{2}-\frac{\delta_t^2}{2}\beta\right)}. \tag{28}$$

Анализ полученных результатов (28) показывает, что в данном случае происходит изменение фазы входного сигнала:

$$arctg\left(\frac{\mathrm{Im}(S_e^-(\omega))}{\mathrm{Re}(S_e^-(\omega))}\right) = \varphi_0 - (\omega-\omega_0)\left(\delta_t^2 - \frac{T_0}{2}\right). \tag{29}$$

При $\omega=\omega_0$ выражение (29) преобразуется к виду (19). Однако, при $\omega\neq\omega_0$, что соответствует смежным с ω_0 частотам, фаза этих частот начинает изменяться в соответствии с зависимостью (29). Также происходит изменение амплитуды спектра в следующее число раз:

$$\Delta A = e^{-\beta\left(\frac{T_0}{2}-\frac{\delta_t^2}{2}\beta\right)}. \tag{30}$$

Таким образом, отсутствие учета реальных физических характеристик РС при осуществлении его спектрального преобразования КФА (STFT) приводит к ухудшению качества получаемого результата. Для сохранения биометрической информативности спектрального преобразования в исследовании предлагается использовать для описания РС его уточненную физическую Гильбертовскую модель.

5. Уточнение цифровых моделей РС

Гильбертовская модель РС [10] базируется на его представлении на вокализованных участках в виде суммы узкополосных составляющих:

$$S(t) = \sum_i s_i(t) = \sum_i A_i(t)\cos(\phi_i(t)), \tag{31}$$

где $A_i(t)$ – амплитуда i-ой узкополосной компоненты, $\varphi_i(t)$ – фаза i-ой узкополосной компоненты.

При этом фаза каждой компоненты определяется следующим образом:

$$\phi_i(t) = arctg\left(\frac{s_i(t)}{s_i^G(t)}\right) + 2\pi k, \tag{32}$$

где s_i^G – сопряженная с $s_i(t)$ функцией в соответствии с моделью Гильберта.

Классический вид синусоидальной модели РС, предложенной МакАуэлем и Куатьери [11], также предполагающей линейную комбинацию синусоид с изменяющимися во времени амплитудами A_k, фазами φ_k и частотами Ω_k, имеет следующий вид:

$$S_{SR}(n) = \sum_{k=1}^{L} A_k \cos(\Omega_k n + \phi_k) \tag{33}$$

Рассматривая только кратные основному тону Ω_0 гармоники, синусоидальная модель принимает вид [10, 11]:

$$S_{SR}(n) = \sum_{k=1}^{L(\Omega_0)} A_k \cos(k\Omega_0 n + \phi_k), \tag{34}$$

где $L(\Omega_0)$ – анализируемое число гармоник РС.

Недостатком применявшейся ранее синусоидальной модели является недостаточная детализация существующих методик определения функций A_k и Ω_k, что затрудняет формализацию их взаимосвязи с индивидуальными биометрическими параметрами, обусловленными физическими характеристиками речеобразующего тракта (Ф2 и Ф3).

В исследовании предложена следующая модифицированная синусоидальная модель $S_{SR}{}^M(n)$ вокализованного фрагмента РС:

$$S_{SR}{}^M(n) = \sum_{k=1}^{L(\Omega_0)} A_k^M(n)\cos(k\Omega_k^M(n)n + \phi_k(n)), \tag{35}$$

где $A_k^M(n)$ – функция, описывающая изменение амплитуды k-ой компоненты (гармоники) в анализируемом окне спектрального анализа;

$\Omega_k^M(n)$ – функция, описывающая изменение частоты k-ой компоненты (гармоники) в анализируемом окне спектрального анализа;

$\varphi_k(n)$ – функция, описывающая изменение фазы k-ой компоненты (гармоники) в анализируемом окне спектрального анализа.

Функция $A_k^M(n)$ описывает огибающую линию и в общем случае может описывать характер изменения амплитуды k-ой гармоники основного тона, включая ее затухание, возрастание, сохранение, либо более сложную временную зависимость с ее первоначальным возрастанием и последующим затуханием. При этом данная функция позволяет также учесть нелинейные эффекты изменения фазы, обусловленные изменением частоты основного тона (ЧОТ) в пределах анализируемого окна.

6. Оценка возможности реконструкции и клонирования речевых вокализмов

Согласно модели (35), на коротких временных интервалах в рамках каждого *R*–шага временного анализа, в качестве первичных описаний РС, представленного в виде суперпозиции элементарных узкополосных процессов или микрогармоник, могут выступать вектора их параметров:

$$\{A_k, \omega_{0k}, \varphi_{0k}\}_{t=rR.} \tag{36}$$

Как показали результаты исследований [12–14], эти основные параметры узкополосных составляющих РС: амплитуды, частоты и фазы, и их треки, далее называемые «следами» фонообъектов, содержатся в динамических спектральных развертках РС или спектрограммах в виде контуров локальных максимумов узкополосных составляющих РС, что особенно характерно для вокализованных участков.

По таким образом найденным или полученным «следам» можно реализовать процесс восстановления гармонической структуры речевых вокализмов с последующим наложением формантной структуры [13, 14] и, тем самым, восстановить РР искаженных РВ. Либо реализовать процесс речевого клонирования или обнаружения факта его применения посредством модернизации и инверсии спектрограмм с выявленными и обработанными следами РВ. Такой спектрально-временной анализ-синтез будем называть трековым или контурным, по аналогии с похожими процедурами цифровой обработки изображений [12]. Отметим важность точного определения следов контуров развития микрогармоник на изображениях сонограмм (чему, впрочем, и посвящена данная работа), позволяющего организовать заявленные и другие приложения в области защиты РИ.

7. Обсуждение результатов экспериментов

Проведенные эксперименты [12–14] подтвердили правильность выбора параметров тестовых речеподобных сигналов (РПС), пригодных для моделирования различных систем защиты РИ (36). Основой формирования таких тестовых сигналов послужило разработанное аналитическое представление РС в виде суперпозиции узкополосных процессов (35) со средней спектральной плотностью, соответствующей усредненному спектру русской речи с учетом рекомендаций [13–15].

Для синтеза тестового РПС в качестве элементарной микрогармоники бралось колебание вида:

$$s(t) = A \cdot \cos(2\pi \cdot (A_s \sin(2\pi f_s t) + f_c)t), \tag{37}$$

т.е. сигнал с несущей f_c амплитуды *A*, промодулированной по частоте гармоническим колебанием с частотой f_s и с девиацией A_s.

Тестовый сигнал представлял собой сумму *n* микрогармоник с девиациями

$$A_j^s = j \cdot \frac{1}{T}, \qquad j = \overline{1,n},$$

где *n* – число микрогармоник, *T* – период основного тона моделируемого речевого сигнала и частотами модуляции

$$f_j^s = \frac{1}{\tau},$$

где τ – длительность слога в моделируемом речевом сигнале.

В качестве несущих принимались $f_j^c = A_j^s / 2$.

В дискретной форме отсчёты сигнала вычислялись по следующей формуле:

$$s_i = \sum_{j=1}^{n} A_j \cos\left(2\pi(f_j^{\,c} + f_j^{\,a})\frac{i}{f_d}\right), \tag{38}$$

где s_i – значение i-го отсчёта; n – число микрогармоник в сигнале; j – номер микрогармоники; A_j – амплитуда j-й микрогармоники; $f_j^{\,c}$ – несущая частота j-й микрогармоники.

Тогда

$$f_j^a = A_j^s \sin(2\pi f_j^s \frac{i}{f_d}) \cdot \left(\frac{0.16 \cdot f_d}{i \cdot f_j^s}\right)$$

– значение отклонения частоты j-й микрогармоники в момент времени $\frac{i}{f_d}$; A_j^s – девиация j-й микрогармоники; f_j^s – частота модуляции j-й микрогармоники.

Результирующий созданный тестовый сигнал (ТС) в частотно-временной области показан на рис. 8. По своей структуре такие синтезированные РПС близки к гласным звукам речи и, по сути, представляют собой их идеализированную модель.

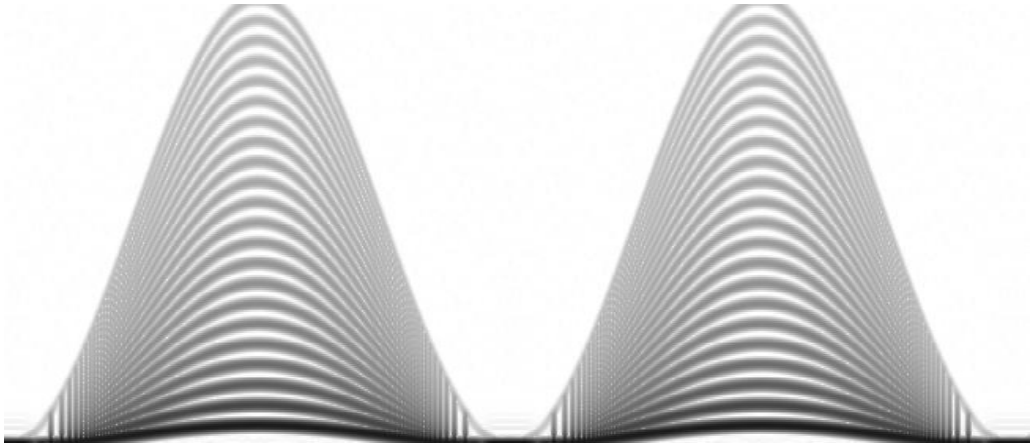


Рис. 8. Тестовый речеподобный сигнал в частотно-временной области

В процессе генерации сигналы подвергались обработке эквалайзером с целью коррекции спектра в соответствии со средним спектром русской речи [14, 15]. Также менялись частота основного тона и длительность «квази» РВ. Спектрограммы строились посредством разработанных авторами звуковых редакторов «Лазурь» и «SoundTools».

Заметим, также, что, помимо искусственных РВ, описываемых выражением (35), в качестве тестового сигнала в приложениях ЗРИ использовались постоянно воспроизводимые образцы фрагментов русской речи с усредненным по [14, 15] спектром. В качестве одного из таких образцов выступала «эталонная усредненная» речь конкретного диктора.

Полученные тестовые речеподобные сигналы микшировались с разными уровнями фоновых акустических шумов. По полученной полезной смеси строилась динамическая узкополосная спектрограмма (рис. 9), на основании визуального анализа изображения которой делался вывод о возможности реконструкции РС и, тем самым, восстановлении его речевой разборчивости (РР) по следам, проступающим на

спектрограмме полезной смеси с заданными параметрами ширины спектральных линий микрогармоник и ЧОТ. Понятно, что чем уже были выявленные на сонограммах треки РВ, тем лучше по ним могли восстанавливать структуру и форму исходного тестового РПС (рис. 8).

Результат одного из таких экспериментов показан на рис. 9. Следы «РВ», хоть и не все как у тестового сигнала, явно обнаруживаются на фоне шумов, что может говорить о незащищенности речевой информации, даже, если рассчитанный по методикам [12–14] показатель (РР) будет ниже порогового значения нормы. При условии, что по таким оставшимся следам квалифицированный злоумышленник, владеющий технологиями цифровой обработки РС, сможет реконструировать исходную начальную гармоническую структуру типа той, что показана на рис. 8 и тем самым восстановить разборчивость искаженной шумами и помехами речи.

Рис. 9. Спектрограмма смеси ТС-«белый шум» SNR=-12 Дб, снятая в одной из точек помещения конфиденциальных переговоров

В [1, 12] отмечено, что проявление параметров речевых вокализмов в виде контуров микрогармоник на фоне спектрально-временного описания полезной смеси составляет основную базу для возможного восстановления злоумышленником защищаемой речевой информации (РИ).

В таких случаях, для обеспечения конфиденциальности переговоров, в защищаемых помещениях необходимо либо понизить уровень тестового РС, либо повысить уровень заграждающего шума или помехи, чтобы следы тестовых РВ не были бы обнаружены на сонограмме.

Заключение

В работе показано, что повышение защищенности РС от подделок и фальсификаций, в том числе, при использовании для этой цели искусственных нейросетей, возможно при учете его характеристик, имеющих индивидуальную окраску, обусловленную уникальностью строения артикуляционной системы человека.

Показано, что восстановление изначальной разрешающей способности спектрального преобразования Фурье возможно при использовании синусоидальной модели РС, учитывающей уникальные для каждого человека биометрические параметры, обусловленные физической природой его речеобразующего тракта

Для сохранения индивидуальных биометрических характеристик РС при спектральной обработке необходимо использовать его уточненную цифровую

физическую модель, основанную на Гильбертовском представлении амплитуды, частоты и фазы узкополосных составляющих участков речевых вокализмов.

Для повышения защищенности речевых технологий, использующих спектральные преобразования, необходимо разработать такое спектральное преобразование, которое способно сохранить потенциальную частотную разрешающую способность, а также передать в составе формируемой спектрограммы уникальные индивидуальные характеристики артикуляционной системы человека.

Разработана модель генерации, выявления и реконструирования тестовых сигналов, аналогичных по частотно-временным характеристикам участкам вокализованной речи, пригодная для использования во многих приложениях защиты речевой информации.

Для реконструкции искаженных участков речевых вокализмов необходимо использовать характеристики оригинальных микрогармоник, выявляемых на изображениях спектрограмм.

В ходе экспериментов получено практическое подтверждение теоретических выкладок о потенциальной возможности восстановления исходного тестового сигнала из смеси «сигнал-шум» с помощью современных методов цифровой обработки сигналов, особенно при наличии базы данных с устной речью диктора, набранной в ходе его открытых выступлений, интервью и т.п. Необходимо учитывать это обстоятельство при построении современных систем защиты речевой информации.

СПИСОК ЛИТЕРАТУРЫ:

1. Дворянкин С.В., Хорев А.А., Козлачков С.Б., Василевская Н.В. Анализ предельных возможностей методов шумопонижения и реконструкции речевых сигналов, маскируемых различными типами помех. Вопросы кибербезопасности. 2024, № 1(59), с. 89–100. URL: https://cyberrus.info/wp-content/uploads/2024/02/vokib-2024-1-st10-s089-100.pdf (дата обращения:13.04.2025).
2. Валорска А.М. Дипфейки и дезинформация. 2020. Кишинев. – ISBN: 978-9975-57-287-3. URL: http://cc.sibimol.bnrm.md/opac/bibliographic_view/745310 (дата обращения:13.04.2025).
3. Дворянкин С.В. Речевая подпись. М.: МТУСИ, 2003. – 238 с.
4. Алюшин А.М., Дворянкин С.В. Анализ перспектив использования фитнес-браслетов в качестве источника биометрической информации при синтезе биоподписи важного документа. Безопасность информационных технологий, [S.l.], т. 31, № 1, с. 63–74, 2024. ISSN 2074-7136. DOI: 10.26583/bit.2024.1.03. – EDN: NUBHAU.
5. Минаев В.А., Дворянкин С.В., Алюшин А.М. Методы биомаркирования защищаемых объектов. Информация и безопасность. 2023, т. 26, вып. 3, с. 321–328. ISSN 1682-7813. DOI: 10.36622/VSTU.2023.26.3.016. – EDN: TGJKIC.
6. Alyushin A. M., Alyushin V. M., Dvoryankin S. V. and Kolobashkina L. V. A Biologically Inspired Approach to Protecting and Verifying the Authenticity of Important Documents. Biologically Inspired Cognitive Architectures 2023 (BICA 2023), v. 1130, p. 50–55. DOI:10.1007/978-3-031-50381-8_7. – EDN: QBOJXR.
7. Любимов Н.А., Захаров Е.В. Математическая модель акустического речеобразования с подвижными стенками речевого тракта. Акустический журнал. 2016, т. 62, № 2, с. 227. DOI: 10.7868/S032079191602009X. – EDN: VLPXHH.
8. Hannukainen A., Lukkari T., Malinen J., Palo P. Vowel formants from the wave equation. I. Acoust. Soc. Am. 2007, v.122, no. 1, p. 1–7. DOI: 10.1121/1.2741599.
9. Женило В.Р. Компьютерная фоноскопия. М.: Из-во Академии МВД РФ, 1995. – 208 с.
10. Коберниченко В.Г. Основы цифровой обработки сигналов: учеб. Пособие. М-во науки и высш. образования Рос. Федерации, Урал. федер. ун-т. Екатеринбург: Изд-во Урал. ун-та, 2018. – 150 с. URL: https://djvu.online/file/YQk6Bw1LOSBXi (дата обращения:13.04.2025).
11. McAulay R. and Quatieri T. Speech analysis/synthesis based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1986, v. 34, no. 4, p. 744–754. DOI: http://dx.doi.org/10.1109/TASSP.1986.1164910.

12. Бонч-Бруевич А.М., Козлачков С.Б. и др. Интерпретация и контурный анализ спектрограмм звуковых сигналов в процессе их шумоочистки. Проблемы информационной безопасности. Компьютерные системы. 2015, № 3, с. 88–99. – EDN: UHSQIF.
13. Дворянкин С.В., Антипенко А.О. Применение фазовых характеристик голосовых вокализмов в решении задач защиты речевой информации. Безопасность информационных технологий, [S.l.], т. 28, № 2, с. 21–33, 2021. ISSN 2074-7136. DOI: 10.26583/bit.2021.2.02. – EDN: JXRJXI.
14. Дворянкин С.В., Макаров Ю.К., Хорев А.А. Обоснование критериев эффективности защиты речевой информации от утечки по техническим каналам. Защита информации. 2007, № 2(14), с. 18–25. – EDN: TRKKQR.
15. Калинцев Ю.К. Разборчивость речи в цифровых вокодерах. М.: Радио и связь. 1991. – 220 с.

REFERENCES:
[1] Dvoryankin S.V., Xorev A.A., Kozlachkov S.B., Vasilevskaya N.V. The analysis of the potential capabilities of methods of noise reduction and reconstruction of acoustic speech signals masked by various types of noise. Voprosy` kiberbezopasnosti. 2024, no. 1(59), p. 89–100. URL: https://cyberrus.info/wp-content/uploads/2024/02/vokib-2024-1-st10-s089-100.pdf (accessed:14.04.2025) (in Russian).
[2] Valorska A.M. Deepfakes and disinformation, 2020, Kishinev. – ISBN: 978-9975-57-287-3. URL: http://cc.sibimol.bnrm.md/opac/bibliographic_view/745310 (accessed:14.04.2025) (in Russian).
[3] Dvoryankin S.V. Speech signature. M.: MTUSI, 2003. – 238 p. (in Russian).
[4] Alyushin A.M.; Dvoryankin S.V. Analysis of the prospects for using fitness bracelets as a source of biometric information when synthesizing the biosignature of an important document. IT Security (Russia), [S.l.], v. 31, no. 1, p. 63–74, 2024. ISSN 2074-7136. DOI: 10.26583/bit.2024.1.03. – EDN: NUBHAU (in Russian).
[5] Minaev V.A., Dvoryankin S.V., Alyushin A.M. Methods of biomarking protected objects. Methods of biomarking protected objects. Informaciya i bezopasnost'. 2023, v. 26, no. 3, p. 321–328. ISSN 1682-7813. DOI: 10.36622/VSTU.2023.26.3.016. – EDN: TGJKIC.
[6] Alyushin A.M., Alyushin V.M., Dvoryankin S.V. and Kolobashkina L.V. A Biologically Inspired Approach to Protecting and Verifying the Authenticity of Important Documents. Biologically Inspired Cognitive Architectures 2023 (BICA 2023), v. 1130, p. 50–55. DOI:10.1007/978-3-031-50381-8_7. – EDN: QBOJXR.
[7] Lyubimov N.A., Zaharov E.V. Mathematical model of acoustic speech production with mobile walls of the vocal tract. Acoustical Physics. 2016, v. 62, no. 2, p. 225–234. DOI: 10.1134/S1063771016020093. – EDN: WSQFGR.
[8] Hannukainen A., Lukkari T., Malinen J., Palo P. Vowel formants from the wave equation. I. Acoust. Soc. Am. 2007, v.122, no. 1, p. 1–7. DOI: 10.1121/1.2741599.
[9] Zhenilo V.R. Computer phonoscopy. Komp`yuternaya fonoskopiya. M.: Iz-vo Akadmii MVD RF., 1995. – 208 p. (in Russian).
[10] Kobernichenko V.G. Fundamentals of Digital Signal Processing: A Textbook. Osnovy` cifrovoj obrabotki signalov: ucheb. posobie; M-vo nauki i vy`ssh. obrazovaniya Ros. Federacii, Ural. feder. un-t. Ekaterinburg: Izd-vo Ural. un-ta, 2018. – 150 p. URL: https://djvu.online/file/YQk6Bw1LOSBXi (accessed: 14.04.2025) (in Russian).
[11] McAulay R. and Quatieri T. Speech analysis/synthesis based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1986, v. 34, no. 4, p. 744–754. DOI: http://dx.doi.org/10.1109/TASSP.1986.1164910.
[12] Bonch-Bruevich A.M., Kozlachkov S.B. i dr. Interpretation and contour analysis of spectrograms of sound signals in the noise reduction process. Problemy` informacionnoj bezopasnosti. Komp`yuterny`e sistemy`. 2015, no 3, p. 88–99. – EDN: UHSQIF (in Russian).
[13] Dvoryankin S.V., Antipenko A.O. Applying the phase characteristics of voice vocalisms in solving problem of protection of speech information. IT Security (Russia), [S.l.], v. 28, no. 2, p. 21–33, 2021. ISSN 2074-7136. DOI: 10.26583/bit.2021.2.02. – EDN: JXRJXI (in Russian).
[14] Dvoryankin S.V., Makarov Yu.K., Xorev A.A. Justification of the criteria for the effectiveness of protecting speech information from leakage through technical channels. Zashhita informacii. 2007, no. 2(14), p. 18–25. – EDN: TRKKQR (in Russian).
[15] Kalincev Yu.K. Speech intelligibility in digital vocoders. Moscow, Radio i svyaz`, 1991. – 220 p. (in Russian).

Статья поступила в редакцию 07.02.2025; одобрена после рецензирования 15.04.202; принята к публикации 20.05.2025
The article was submitted 07.02.2025; approved after reviewing 15.04.202; accepted for publication 20.05.2025