Сергей В. Дворянкин[1], Никита С. Дворянкин[2], Роман А. Устинов[3]
[1,3]Финансовый университет при Правительстве Российской Федерации,
Ленинградский пр-т, 49, г. Москва, 125993, Россия
e-mail: SVDvoryankin@fa.ru, http://orcid.org/0000-0001-6908-0676
e-mail: public-ura@yandex.ru, http://orcid.org/0000-0002-8454-9951
[2]Национальный исследовательский ядерный университет «МИФИ»,
Каширское ш., 31, г. Москва, 115409, Россия
e-mail: nik.dvrn@gmail.com, http://orcid.org/0000-0002-1580-7179

РАЗВИТИЕ ТЕХНОЛОГИЙ ОБРАЗНОГО АНАЛИЗА-СИНТЕЗА АКУСТИЧЕСКОЙ (РЕЧЕВОЙ) ИНФОРМАЦИИ В СИСТЕМАХ УПРАВЛЕНИЯ, БЕЗОПАСНОСТИ И СВЯЗИ

DOI: http://dx.doi.org/10.26583/bit.2019.1.07

Аннотация. Голосовые коммуникации были и остаются одним из основных способов человеческого общения и человеко-машинного обмена. Цель работы состоит в создании инструментария моделирования эффективных механизмов речепреобразования, формирования речеподобных сигналов с заданными свойствами для построения новых эффективных систем обработки и защиты акустической (речевой) информации. Методы. В работе предложен оригинальный подход, заключающийся в преобразовании полутонового изображения узкополосной спектрограммы речевого сигнала в бинарное, необходимой трансформации полученного образа под решаемую задачу с возможностью обратного перехода к полутоновому спектральному изображению с последующим синтезом по нему нового речеподобного сигнала с нужными характеристиками. Результаты. Совершенствование модели речеобразования, использование свойств слухового восприятия и учета особенностей формирования бинарных спектрограмм позволяют существенно сократить объем акустической (речевой) информации, содержащейся в полутоновых спектральных описаниях, без потери смыслового содержания и узнаваемости, предоставляет возможность использования богатого и хорошо апробированного арсенала способов распознавания и цифровой обработки бинарных и полутоновых изображений для решения задач обработки и защиты речевой информации. Выводы. В работе оценены возможности и перспективы использования образного анализа-синтеза применительно к графическим образам узкополосных сонограмм и изображениям иного рода в решении задач аудио стеганографии, цифровой шумоочистки и реконструкции искаженных фонограмм, аудиомаркирования значимой информации, сжатия-восстановления речи.
Ключевые слова: защита речевой информации, образный анализ-синтез, речевой сигнал, бинаризация изображений спектрограмм, кратковременное преобразование Фурье.
Для цитирования: ДВОРЯНКИН, Сергей В.; ДВОРЯНКИН, Никита С.; УСТИНОВ, Роман А. РАЗВИТИЕ ТЕХНОЛОГИЙ ОБРАЗНОГО АНАЛИЗА-СИНТЕЗА АКУСТИЧЕСКОЙ (РЕЧЕВОЙ) ИНФОРМАЦИИ В СИСТЕМАХ УПРАВЛЕНИЯ, БЕЗОПАСНОСТИ И СВЯЗИ. Безопасность информационных технологий, [S.l.], p. 64-76, 2019. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/1186>. Дата доступа: 19 feb. 2019. doi:http://dx.doi.org/10.26583/bit.2019.1.07.

Sergey V. Dvoryankin[1], Nikita S. Dvoryankin[2], Roman A. Ustinov[3]
[1,3]Financial University under Government of the Russian Federation
Leningradsky Prospekt, 49, Moscow, 125993, Russia
e-mail: SVDvoryankin@fa.ru, http://orcid.org/0000-0001-6908-0676
e-mail: public-ura@yandex.ru, http://orcid.org/0000-0002-8454-9951
[2]National Nuclear Research University MEPHI,
Kashirskoe shosse, 31, Moscow, 115409, Russia
e-mail: nik.dvrn@gmail.com, http://orcid.org/0000-0002-1580-7179

Improvement of image analysis/synthesis technologies of acoustic (speech) information for the control, safety and communication systems

DOI: http://dx.doi.org/10.26583/bit.2019.1.07

Abstract. Voice communications have been and remain one of the main ways of human communication and human-machine exchange. The aim of the work is to create a tool for modeling effective mechanisms

of speech transformation, the formation of speech-like signals with desired properties for the construction of new effective systems of processing and protection of acoustic (speech) information. Methods. The paper proposes an original approach consisting in the conversion of a halftone image of a narrow-band spectrogram of a speech signal into a binary one, the necessary transformation of the resulting image for the problem to be solved, with the possibility of a reverse transition to a halftone spectral image, followed by the synthesis of a new speech-like signal with the desired characteristics. Results. Improving the model of speech formation, using the properties of auditory perception and taking into account the features of the formation of binary spectrograms can significantly reduce the volume of acoustic (speech) information contained in the halftone spectral descriptions, without loss of semantic content and recognition, provides an opportunity to use a rich and well-tested arsenal of methods of recognition and digital processing of binary and halftone images to solve problems of processing and protection of speech information. Summary. In the work the possibility and perspectives of use of the visual analysis-synthesis as applied to the graphic images of the narrow-band sonograms and images of a different kind in the task of audio steganography, digital noise reduction and reconstruction of distorted phonograms, audiomarkering significant information, compression-restoration of speech.

Keywords: information protection, figurative analysis-synthesis, speech signal, binarization of speech spectrum images, short-time Fourier transform.

For citation: DVORYANKIN, V. Dvoryankin V.; DVORYANKIN, Nikita S.; USTINOV, Roman A. Improvement of image analysis/synthesis technologies of acoustic (speech) information for the control, safety and communication systems. IT Security (Russia), [S.l.], p. 64-76, 2019. ISSN 2074-7136. Available at: <https://bit.mephi.ru/index.php/bit/article/view/1186>. Date accessed: 19 feb. 2019. doi:http://dx.doi.org/10.26583/bit.2019.1.07.

Введение

На сегодняшний день существует достаточно широкий набор методов и средств обработки и защиты значимой акустической (речевой) информации от НСД. В зависимости от требований безопасности и условий применимости используют тот или иной способ или их комбинации [1].

Отдельного внимания исследователей заслуживает технология образного анализа-синтеза (ОАС) акустических речевых сигналов (АРС) как оригинальная база моделирования существующих способов защиты и обработки акустической речевой информации (АРИ) с возможностью оценки эффективности их работы, так и как платформа создания и использования ранее не применявшихся способов речепреобразования для их реализации в различных перспективных системах голосового управления, безопасности и связи, в которых циркулирует АРИ [2 – 4].

Суть указанной технологии ОАС или по-другому технологии «звук-изображение-звук» состоит в преобразовании звукового или речевого сигнала в изображение узкополосной спектрограммы с применением к нему развитых эффективных методов цифровой обработки изображений с последующим переходом от него к новой волновой форме звукового или речеподобного сигнала (РПС) с требуемыми характеристиками [2]. Понятно, что в таком случае на тех же базовых элементах и принципах можно реализовать «зеркальное» преобразование «изображение-звук-изображение», которое тоже может найти свое применение в указанных системах.

Исследованию возможностей ОАС изображений спектрограмм АРС в различных областях применения: скремблирование, аудио стеганографии, кодировании и компрессии речи, нейтрализации помех и искажений, идентификации говорящего и т.п. посвящено множество работ [2 – 4], в которых описаны различные варианты успешного применения технологии ОАС за счет реализации алгоритмов синтеза звуковых сигналов с требуемыми свойствами на основе заданного изображения спектрограммы. В качестве базовых в исследованиях использовались полутоновые (в уровнях серого цвета) изображения с достаточной информационной избыточностью.

В данной работе сделана попытка сокращения информационной избыточности образных описаний речевых сигналов (РС) без потери разборчивости и узнаваемости за счет выделения и использования значимых (главных) контуров или компонент, перехода от полутоновых к бинарным изображениям без потери информативности исходного РС

для последующего применения к ним проработанного аппарата цифровой обработки бинарных изображений [5, 6].

Оценены возможности прямого и зеркального ОАС применительно к решению задач акустической стеганографии, цифровой шумоочистки и реконструкции искаженных фонограмм, аудиомаркирования значимой информации, сжатия-восстановления речи.

Предполагается, что применение бинарного, прямого и зеркального ОАС ускорит процессы моделирования существующих и перспективных звуко- и речепреобразующих устройств, найдет или расширит области их применения в системах безопасности, управления и связи.

1. Синусоидальная узкополосная модель речеобразования

В основе технологии ОАС лежит синусоидальная модель речевого сигнала, которая характеризуется амплитудами, частотами и фазами L узкополосных составляющих или синусоидальных волн, составляющих РС.

Эти параметры оцениваются на этапе анализа по кратковременному преобразованию Фурье (КПФ) с базой N алгоритма его быстрой реализации БПФ, с использованием простого алгоритма пикового подбора [2, 7] на каждом шаге наблюдения. Данное преобразование используется всякий раз для вычисления амплитудного, частотного и фазового содержания синусоидальных составляющих по локальным спектральным срезам (сечениям) исходного акустического речевого сигнала по мере его изменения с течением времени.

На практике процедура вычисления КПФ состоит в том, чтобы разделить сигнал длительного времени на более короткие, перекрывающиеся сегменты равной длины, а затем вычислить через быстрое преобразование Фурье амплитудный и фазовый спектры, отдельно на каждом коротком сегменте (фрейме). Ансамбль получаемых амплитудных спектральных срезов по времени позволяет строить изображения спектрально-временных разверток, изображений спектрограмм, с возможным применением к последним всего арсенала средств цифровой обработки изображений [5, 6], в том числе для выполнения базовой операции выделения пиковых треков и амплитудно-фазово-частотных параметров L узкополосных или синусоидальных составляющих исходного РС.

Исходный РС можно рассматривать как выход линейной системы, представляющей характеристики речевого тракта при поступлении на нее сигнала возбуждения от голосовых связок. Согласно такому представлению процесса речеобразования формируемый РС при длительности анализируемого фрейма речи до *40 мс* (оптимально *6-8 мс*) [2, 7] может быть представлен как:

$$s(n) = \sum_{i=1}^{L} A_i e^{\frac{-n^2}{2\sigma}} \cos(\omega_i n + \varphi_i) + e(n) \quad (1)$$

где: n – номер временного отсчета; L – количество значимых синусоид; A_i – амплитуда i-й синусоиды; ω_i – частота i-й синусоиды; φ_i – фаза i-й синусоиды, σ – эффективная ширина окна функции Гаусса, $e(n)$ – остаточный сигнал.

В таком виде исходный речевой сигнал можно рассматривать как суперпозицию узкополосных сигналов или вейвлетов Морле. Такое представление (1) можно распространить и на другие акустические сигналы.

Изменения в спектральных компонентах треков узкополосных (синусоидальных) составляющих (УС) отслеживаются на спектрограмме с использованием понятий «рождение» и «смерть», лежащих в основе принятого синусоидального представления [7].

На этапе синтеза для каждого трека (следа) или контура УС, определенного на изображении спектрограммы (например, траектории линий гармоник основного тона на вокализованных фреймах), к заданным параметрам частоты и амплитуды трека будет присоединяться фазовая функция, необходимая для разворачивания и интерполяции фазы и построенная таким образом, чтобы фазовый след был максимально гладким [2].

В зависимости от решаемой задачи выбирается либо исходная, вычисленная по комплексному спектру исходного сигнала, либо или искусственная, вычисленная на основе анализа изображения амплитудной спектрограммы, или, как увидим позже, изображения иного рода, фазовая функция, которая и применяется для синтеза нового речеподобного сигнала. Синтез может проходить либо в блоке обратного КПФ, либо в гребенке синусоидальных генераторов, выход каждого звена которой модулируется амплитудой на частоте, найденной УС и добавляется к другим синусоидальным волнам, чтобы сформировать окончательный вывод речи или звука, синтезированных по изображению спектрограммы и изображениям иного рода.

Полученная по исходной спектрограмме синтетическая форма волны РС совпадает с исходной волной в случае использования оригинальной фазы треков и сохраняет общую форму огибающей волны исходного сигнала для искусственно подобранной фазы. Звучание (разборчивость, узнаваемость, естественность, громкость и т.д.) синтезированных РС для обоих типов выбираемой фазы практически не отличается по звучанию от оригинальной речи, звука. При наличии шума сохраняются перцептивные характеристики речи, а также и сам шум [2, 7].

Заметим, что снижение в принятой модели РС числа синусоидальных компонент с *N/2* (где *N* база КПФ на основе алгоритма БПФ, как правило, выбирается равной – *1024* для частоты дискретизации *8-10 КГц*) до *L* (максимальное число обычно выбирается равным *64* [2] или *80* [7]) никак не сказывается на качественных (громкость, естественность, узнаваемость) и смысловых (разборчивость, эмоциональность) характеристиках речи, звука, синтезируемых по амплитудно-частотно-фазовым параметрам пикового подбора треков узкополосных составляющих РС на изображениях спектрограмм без каких-либо изменений последних.

Дальнейшее сокращение числа синусоидальных компонент возможно за счет использования свойств слухового восприятия и особенностей восприятия изображений спектрограмм.

2. Образный анализ-синтез речи

Система анализа-синтеза речи с использованием свойств слухового восприятия (учета работы органов слуха, эффектов частотного и временного маскирования, психоакустики и др.) и контурного анализа узкополосных спектрограмм представлена на рис. 1 [2, 8].

В блоке анализа входной дискретизированный речевой сигнал *S(n),* подвергаясь КПФ, периодически от фрейма к фрейму взвешивается временным окном *W(n)* (например, усеченным окном Гаусса или окном Хэмминга), вычисляется его спектр $|S(f)|$. На каждом спектральном срезе $|S(f)|$ осуществляется отбор пиков главных синусоидальных компонент, наиболее подходящих с точки зрения наилучшего перцептуального качества синтезированной речи, её слухового восприятия. На получаемом по развертке спектральных срезов изображении спектрограммы отрисовываются треки или контуры этих пиков для каждой УС.

В блоке синтеза по положениям контуров выбранных пиков определяются частоты наиболее мощных синусоид, определяющих основное звучание РС, и их амплитуды. Оригинальные фазы синусоид определяются по действительной и мнимой компонентам спектра *S(f)* на соответствующих найденных частотах [7] или вычисляются искусственным путем [2] по функции развития фазы, определяемой по изображению (разверткам) амплитудного спектра $|S(f)|$ или изображению иного содержания.

Процесс синтеза новых речеподобных сигналов с заданными свойствами сводится к обратному КПФ изменённого в зависимости от решаемой задачи спектрального среза с заданным пиковым подбором главных УС или суммированию сгенерированных главных синусоидальных компонент с найденными для них в процессе анализа амплитудами, фазами и частотами (рис. 1).

При этом для получения в процессе синтеза приемлемого качества речи необходимо генерировать синусоиды, непрерывно изменяющиеся во времени. С этой целью применяется частотное упорядочивание синусоид и интерполяция их параметров от фрейма к фрейму [2, 7, 9].

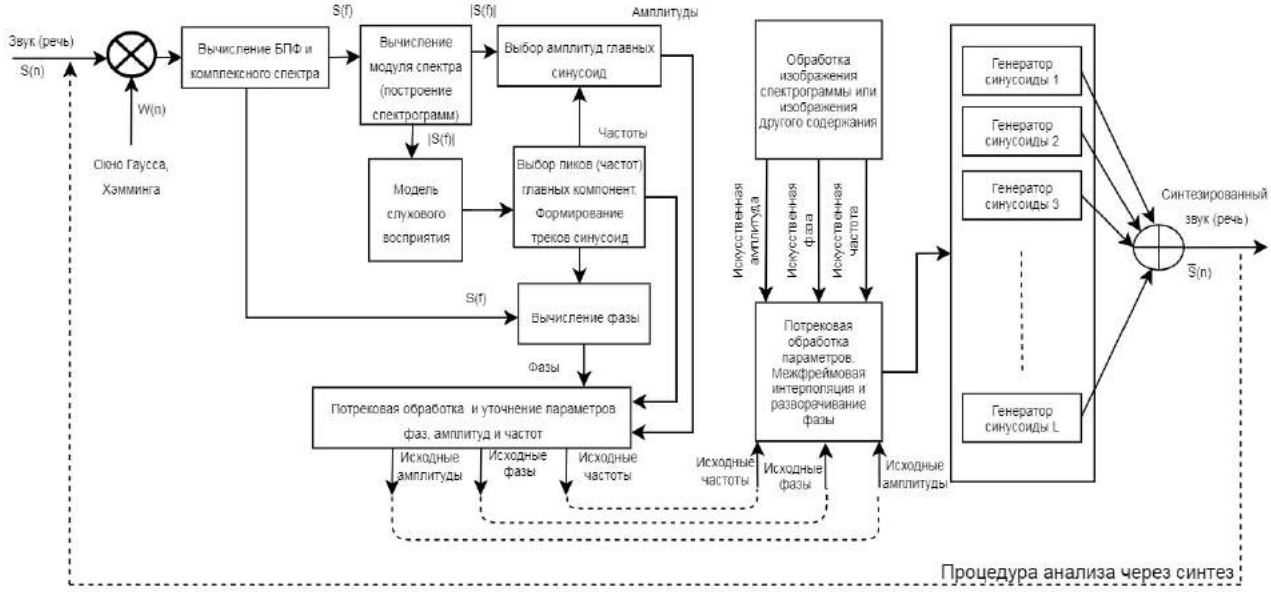


Рис.1. Система образного анализа-синтеза речевого сигнала на основе синусоидальной узкополосной модели и свойств слуха
(Fig.1. Image analysis and speech signal synthesis system based on a sinusoidal narrowband model and hearing properties)

Система образного анализа-синтеза акустического (речевого) сигнала на основе синусоидальной узкополосной модели РС и учета свойств слуха, представленная на рис. 1, послужила основой для моделирования работы различных речепреобразующих устройств [2], когда на выходе анализирующей части системы (рис. 1) собирался ансамбль спектральных срезов модуля спектра (спектральные развертки), рассматриваемый в дальнейшем как изображение. На получаемых графических образах РС определялись треки (линии) пиков локальных максимумов, по амплитудно-частотно-фазовым параметрам которых в блоке синтеза (рис. 1) генерировался новый речеподобный сигнал в соответствии с заданным образом модифицированной спектрограммой.

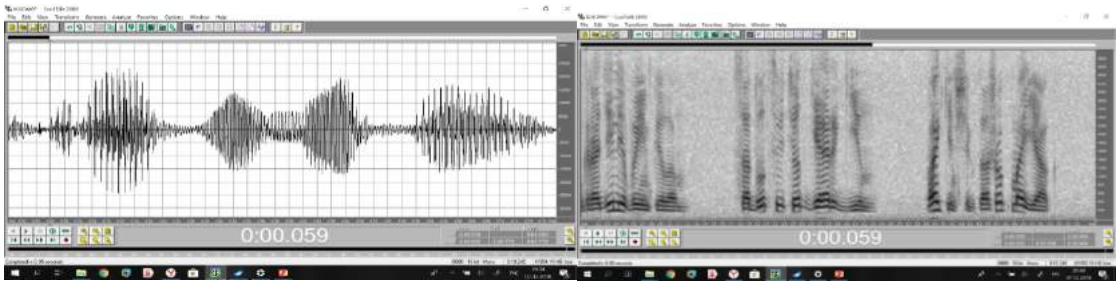
Процесс анализа-синтеза звуков и речи, на основе принятой модели речеобразования *(1),* был промоделирован и доказал свою адекватность. При обработке РС использовалось временное окно Хемминга длительностью около *24 мс* с перекрытием *20 мс* при частоте дискретизации сигнала *10 кГц* [7] и усеченное окно Гаусса с эффективной шириной *16 мс* с перекрытием *6 мс* при частоте дискретизации сигнала *8 кГц* [2]. База быстрого преобразования Фурье составляла *N=512* [7] и *N=1024* отсчётов [2].

Волновые формы исходного РС, а также сигналов, синтезированных в системе ОАС с параметрами из работы [2] по трекам локальных максимумов УС, найденных на изображениях спектрограмм, с оригинальной и искусственной фазой на синтезе показаны на рис. 2.

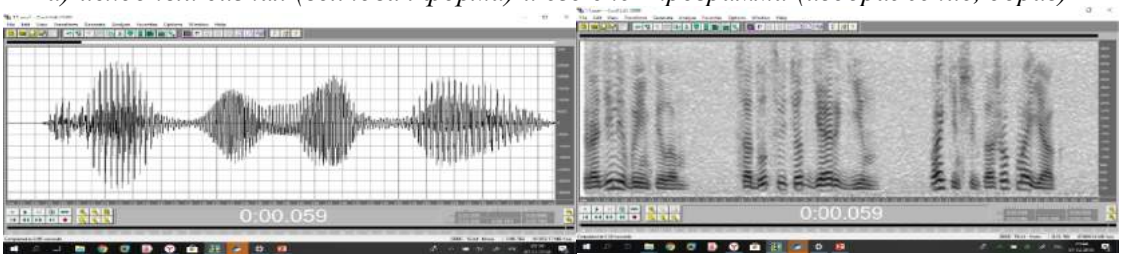
Синтез проводился для всех найденных на изображениях спектрограмм треков локальных максимумов УС. Звучание исходного и синтезированных РС практически совпадает. Спектрограммы исходного и двух новых синтезированных речеподобных сигналов также практически совпадают.

Процедуры ОАС в соответствии с рис. 1 удобно использовать для моделирования систем синхронной и асинхронной двухканальной шумоочистки, когда в блоке анализа одновременно формируются изображения спектрограмм полезной смеси и помехи (шума) с последующим выделением на них треков с амплитудно-частотно-фазовыми

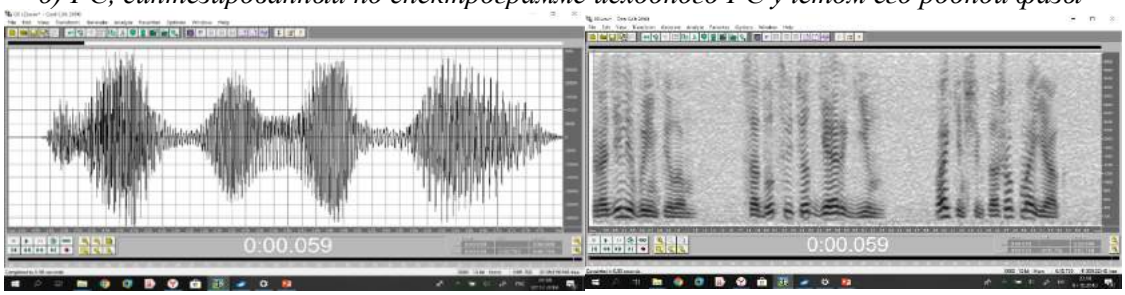
параметрами синусоид. Далее происходит совмещение треков обеих спектрограмм, чтобы в результате потрековой обработки на спектрограмме полезной смеси остались только треки (следы) речи, а треки, совпавшие с треками (следами) на помеховой спектрограмме, затираются. По оставшимся полезным трекам синтезируется очищенный речевой сигнал. Эксперименты показали высокую эффективность такой процедуры шумоочистки при отношениях сигнал/шум *-25 Дб* для узкополосных помех и *-12÷-18 Дб* для сложных шумовых и речеподобных помех. При использовании голосовой базы данных диктора и процедур реконструкции и восстановления изображений с использованием образцов эффективность шумоочистки речи через шумоочистку изображений спектрограмм может ещё более повыситься.



а) исходный сигнал (волновая форма) и его спектрограмма (изображение, образ)



б) РС, синтезированный по спектрограмме исходного РС учетом его родной фазы



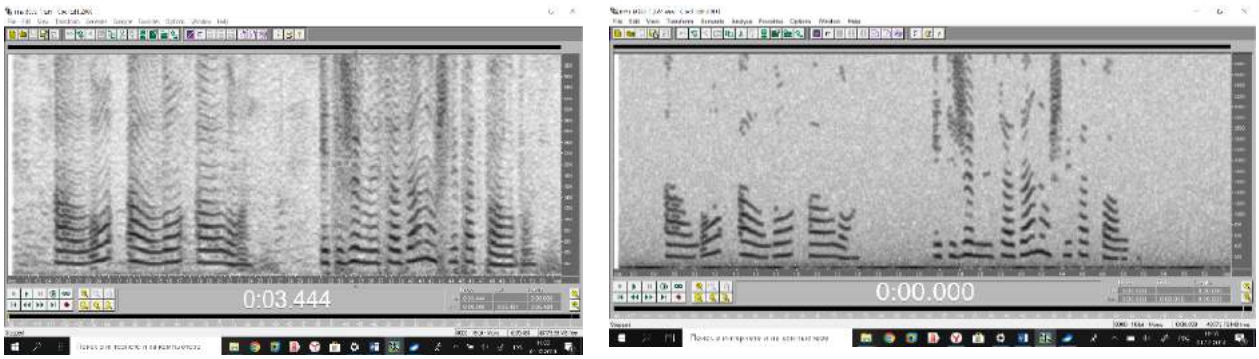
в) сигнал, синтезированный по спектрограмме исходного РС с подставкой искусственной фазы

Рис. 2. Осциллограммы и спектрограммы исходного РС (а) и сигналов, синтезированных по спектрограммам с исходной (б) и искусственной фазой (в)
(Fig. 2. Oscillograms and spectrograms of the original speech (a) and signals synthesized by spectrograms with the original (b) and with the artificial phase (c).)

В качестве следующего шага усовершенствования и развития технологии ОАС предлагается использовать изображение не всего спектра исходного РС, а только той его части, в которой содержится основная энергетическая составляющая РС (главные компоненты), отвечающая за его разборчивость. Результаты экспериментов по определению на изображениях спектрограмм главных синусоидальных компонент и синтезу по ним нового сигнала представлены на рис. 3.

При этом, как уже отмечалось, построение и анализ полутоновых изображений спектрограмм РС на основе КПФ (рис. 3а) производится в анализирующей части системы ОАС. Заметим, что по оси абсцисс задается время, по оси ординат – частота. В уровнях серого цвета – мощность на данной частоте в данное время. Максимальная – чёрный цвет, минимальная – белый.

По получаемой или задаваемой спектрограмме определяются амплитудно-фазово-частотные параметры треков локальных максимумов (пиков), по исходным или измененным значениям, которых (в зависимости от решаемой задачи) с использованием синусоидальной модели производится синтез нового речеподобного сигнала (рис. 1).



а) б)

Рис. 3. Спектрограммы исходного РС (а) и сигнала, синтезированного по трекам главных компонент (линиям протяжки пиков локальных максимумов) на исходной спектрограмме (б)
(Fig. 3. Spectrograms of the original speech (a) and the signal synthesized along the tracks of the main components (lines of broach of the peaks of local maximum) in the original spectrogram (b))

Если, например, сохраняются параметры всех УС, составляющих исходный сигнал, то изображения и звучание исходного и синтезируемого сигналов совпадают (рис. 2). Если при синтезе учитываются только главные компоненты, то изображения спектрограмм исходного и нового просинтезированного РС могут отличаться (рис. 3), тем не менее звучание исходного и синтезированного по главным компонентам сигналов останется схожим.

В проведенных экспериментах число главных синусоидальных компонент на каждом рассчитанном спектральном срезе выбиралось следующим образом:

– совместная энергетическая мощность оставленных главных компонент должна быть не менее 80 % от общей мощности исходного среза;

– главные синусоиды должны размещаться на частотно-временной сетке в зоне присутствия как минимум 2-х разных формант.

В результате моделирования было определено, что минимальное число L таких главных синусоидальных компонент, выявляемых на изображениях узкополосных спектрограмм, в зависимости от типа голоса (мужской, женский детский) и условий акустического фона (шумы, помехи) может составлять от 4-х до 16-ти (рис. 3) как на вокализованных, так и на невокализованных участках.

Еще раз отметим, что при таком подходе к выбору качество звучания исходного РС и синтезированного по найденным на исходной спектрограмме главным компонентам либо с оригинальной, либо с искусственной фазой практически совпадает.

Заметим, что в ряде задач защиты и обработки АРС после синтеза новых сигналов по главным УС для улучшения комфортности звучания рекомендуется добавлять (подмешивать) фоновый розовый шум, а главные гармоники в области верхних частот (формант) подчеркивать усилением. Для выделения главных УС в анализирующей части системы ОАС (рис. 1) дополнительно к модели слуха [8, 9] можно также использовать процедуры учета временного и частотного маскирования, а также психоакустики [10], доказавшие свою эффективность, например, в сжатии-восстановлении АС (РС) по стандарту MP3.

Определив, таким образом, главные компоненты на полутоновых изображениях спектрограмм, можно перейти к их бинаризации, которая в еще большей степени сократит информационную избыточность спектральных описаний АС (РС) при сохранении их общей информативности.

3. Бинаризация полутоновых спектрограмм речевого сигнала

В качестве очередного шага совершенствования и развития технологии ОАС предлагается использовать изображение не всего спектра исходного РС, а только той его части, в которой содержится основная энергетическая составляющая РС (главные компоненты), отвечающая за его разборчивость, звучание, а само изображение целесообразно преобразовать из полутонового в черно-белое (бинарное), как это поэтапно показано на рис. 4 [11, 12].

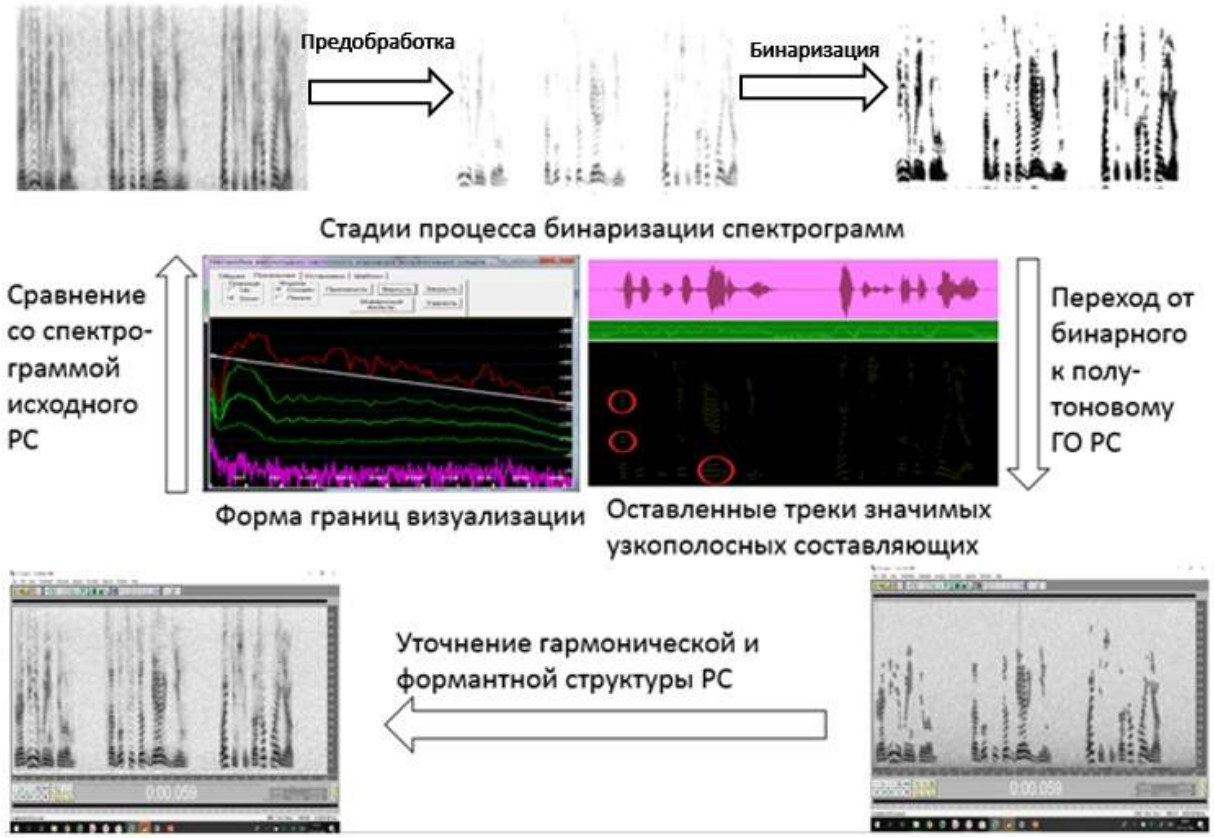


Рис. 4. Технология бинаризации изображения спектра речевого сигнала
(Fig. 4. Binarization technology of speech spectrum images)

В начале на основе КПФ строится и анализируется исходное полутоновое изображение спектрограммы РС. Отметим, что по оси абсцисс – время, ординат – частота. В уровнях серого цвета – мощность на данной частоте в данное время. Максимальная – чёрный цвет, минимальная – белый.

Затем в процессе предобработки построенной спектрограммы происходит выделение треков (следов) и параметров (частот, амплитуд и фаз) главных синусоидальных компонент.

Следующий шаг к улучшению образных описаний РС – бинаризация, которая предполагает корректное отражение найденных в процессе предобработки параметров главных синусоид в новом бинарном представлении. На бинарной спектрограмме треки найденных и оставленных главных синусоид сохраняются, но изменяется их толщина, которая ставится в соответствие с изменяемой во времени амплитудой главных компонент по логарифмическому закону.

Заметим, что процесс бинаризации является обратимым. От рассчитанного бинарного изображения с помощью обратных преобразований можно перейти к полутоновому представлению спектрограммы, практически совпадающему с исходным.

Пример созданной бинарной сонограммы речевой подписи показан на рис. 5.

Бинарная сонограмма или речевая подпись (РП), может быть отсканирована, сфотографирована, переведена в полутоновой образ и практически мгновенно преобразована в звук с помощью специально разработанного приложения для смартфона, планшета или ПК [13]. Синтетическая речь 6-ти секундной длительности, телефонного качества звучания, практически неотличима от исходной.

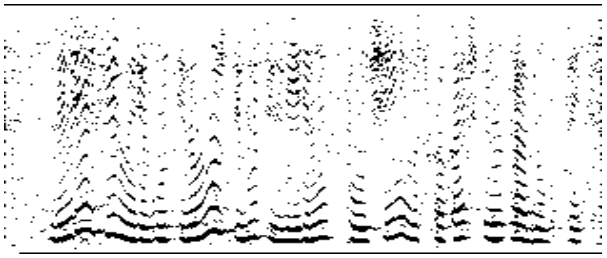


Рис. 5. Пример бинарной сонограммы: 6 секунд телефонного качества
(Fig. 5 Example of binary sonogram: 6 seconds of phone quality)

Такую речевую подпись можно использовать в приложениях компьютерной стеганографии, размещая ее в фото и видеоматериалах с целью охраны их авторства с использованием положений и результатов работ [14, 15, 16]. Можно, как и обычную подпись, размещать РП на бумажных документах с целью их защиты от фальсификации и подделки. В отличие от других элементов защиты, защищающих сам носитель информации, речевая подпись подобно электронной связана с информационным содержанием документа. Замятие бумаги и нанесение пятен не исказит содержания РП и выразится при синтезе в виде шорохов и потрескиваний как в телефонной трубке при плохой связи.

На базе рассмотренных бинарных описаний можно создавать новые алгоритмы скремблирования. Также интересна оценка возможности сжатия-восстановления речи через сжатие-восстановление бинарных сонограмм. Расчеты показывают возможность создания широкодиапазонного помехоустойчивого аудиокодека, работающего на скоростях от 1 до 64 Кбод с плавной адаптацией к пропускной способности канала речевой связи. С учетом использования голосовой базы данных конкретного диктора можно достичь 400 – 600 бит/с, приближаясь вплотную к теоретическому пределу – 70 бит/с.

Варианты решения других задач защиты и обработки АРИ посредством обработки бинарного изображения спектрограмм в моделируемых на основе ОАС речепреобразующих устройствах также могут оказаться предпочтительней по сравнению с известными. Это обусловлено и тем, что алгоритмы обработки бинарных изображений наиболее просты в реализации и требуют меньших вычислительных мощностей [5].

Использование бинарных изображений АРС позволит воспользоваться уже накопленным значительным потенциалом обработки бинарных изображений в разных сферах и в итоге не только усилить защищенность значимой РИ от возможного НСД и помех, но и расширить области применения бинарных образов АРС для решения большого круга задач защиты и обработки РИ.

Отметим, что к получаемому бинарному изображению узкополосных сонограмм уже могут быть применены алгоритмы цифровой обработки черно-белых бинарных изображений, что может быть весьма удобным при решении задач распознавания, сжатия и защиты речевой информации и др.

4. «Зеркальное» (инверсное) преобразование «изображение-звук-изображение»

На основе выше рассмотренных базовых элементах и принципах прямого ОАС можно реализовать его «зеркальное» преобразование: «изображение-звук-изображение». Тогда изображение любого вида и содержания может быть преобразовано в звуковой

сигнал, спектрограмма которого будет похожа, например, на исходное фото, как это показано на рис. 6.

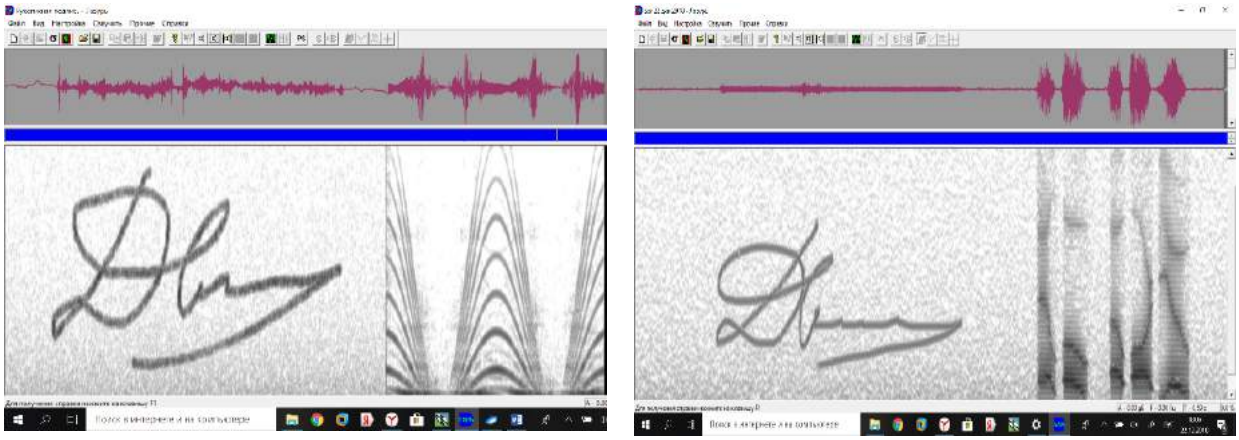
Звуковые сигналы со спектральными характеристиками в виде различных фотоснимков могут найти своё применение в виде специфических аудиомаркеров, которыми можно, например, подтверждать, подлинность голосовых команд в системах речевого управления или доверенность среды передачи голосовых сообщений. Аудиомаркеры можно использовать в системах голосовой аутентификации используя речеподобный сигнал с заданным фото, в качестве парольной фразы в процессе непрерываемого голосового контакта с поставщиками необходимых услуг или call-центром.

а) б)

Рис. 6. Фотопортрет - а) и его реализация в виде спектральной развертки осциллограммы звука-носителя – б)

(Fig. 6 Photo - a), and its implementation in the form of a spectral scan of the sound carrier oscillogram – b))

Еще более интересные возможности открывает создание и применение в системах дистанционной аутентификации пользователя защищенного информационного ресурса речеподобного образа его рукописной подписи (рис. 7).



а) б)

Рис. 7. Представление рукописной подписи в качестве спектрограммы речеподобного сигнала, его статических и динамических характеристик: а) в частоте основного тона; б) в количестве и направленности формант

(Fig. 7. Sonification of the handwritten signature as a spectrogram, static and dynamic characteristics of the speech-like signal: a) in the frequency of the pitch tone; b) in the quality and direction of the formants)

Причем в этом случае для сравнения с эталоном можно использовать не только статическое изображение самой рукописной подписи в спектре звука-носителя, но и динамические характеристики ее написания, сохраненные в виде изменений параметров частоты основного тона и количества формант в дополнительно создаваемом образе речеподобного сигнала, соответствующем движению пера (пальца) на экране смартфона.

Так, движение пера (пальца) на экране смартфона «вверх-вниз» изменяет частоту основного тона на спектрограмме синтезируемого речеподобного сигнала (РПС) от *100* до *250 Гц*. Движение «слева-направо» соответствует двум формантам в спектре РПС с глобальными максимумами в области *200* и *500 Гц*. Обратное движение – «справа-налево» характеризуется добавлением в спектральный образ РПС третьей форманты с глобальным максимумом в *1500 Гц*.

Силе нажима на кончик пера в параметрах речеподобного сигнала будет соответствовать движение формант на спектрограмме РПС также сверху-вниз или снизу-вверх.

Такая техника сонификации рукописной подписи как типовой частный случай применения ОАС проста в реализации и весьма перспективна.

Наличие для взаимного сравнения трёх образов рукописной подписи: один, как эталон-образец создается при первоначальном личном присутствии пользователя в службе безопасности и обслуживания, а два других определяются из статических и динамических характеристик спектра РПС, передаваемого по каналу связи, существенно повышают надежность аутентификации удаленного пользователя защищенного информационного ресурса.

РПС с образцами подписей также может рассматриваться как некая парольная фраза, и тогда можно ещё усилить надежность аутентификации, добавив к существующим хорошо отработанным процедурам сравнения рукописных подписей такую же хорошо отработанную процедуру голосового паролирования от известных российских и зарубежных производителей.

Возможны и другие не менее интересные приложения инверсного преобразования «изображение-звук-изображение». Например, совмещение в системах удаленной аутентификации двух выше описанных приложений инверсного ОАС.

Заключение

Работа посвящена развитию хорошо зарекомендовавшей себя в решении задач защиты и обработки речевой информации технологии образного анализа-синтеза АРС.

В рамках указанной технологии исследован подход, заключающийся в преобразовании полутонового изображения АРС в бинарное, его модификации в соответствии с решаемой задачей, с возможностью обратного перехода к полутоновому с последующим синтезом по нему нового речеподобного сигнала с заданными свойствами.

Использование такого подхода существенно сокращает объем представляемой информации, даёт возможность использования богатого и хорошо апробированного арсенала методов обработки бинарных черно-белых изображений в задачах обработки и защиты акустической речевой информации.

Для того, чтобы по бинарному изображению сонограммы можно было восстановить РС с сохранением таких его важных характеристик как звучание, узнаваемость или разборчивость, должен выполняться ряд определенных в данном исследовании условий:

1. Необходимо осуществлять предобработку исходного полутонового изображения спектрограммы, подвергающегося бинаризации (достаточно оставить примерно 80 % энергетической составляющей спектрального среза, что устранит информационную избыточность изображения, но при этом не пострадают важные характеристики РС).
2. После предобработки на изображении спектрограмм вокализованных участков в зоне глобальных спектральных максимумов (формант) должно оставаться не менее двух рядом стоящих гармоник РС.
3. Границы коридора визуализации треков узкополосных составляющих РС на исходной полутоновой спектрограмме должны иметь параметры, обеспечивающие выполнение условий 1, 2.

Помимо выигрыша в объеме представляемой информации и использовании простых и хорошо зарекомендовавших себя алгоритмов распознавания и обработки черно-белых изображений, бинарные спектральные образы можно использовать в качестве помехоустойчивой речевой подписи для защиты бумажных и электронных документов от фальсификации и подделки, причем такие аудиомаркеры можно размещать на текстовом документе и в неявном виде, используя методы стеганографии.

В традиционных приложениях компьютерной стеганографии бинарные изображения могут также оказаться предпочтительнее, поскольку их, например, гораздо проще встраивать в стегоконтейнеры по сравнению с полутоновыми изображениями [14, 15, 16]. Бинарные образы РС могут быть использованы в виде оригинальных цифровых водяных знаков для охраны авторских прав на аудиовизуальную и печатную продукцию.

На основе бинарного ОАС, реализуя сжатие речи через сжатие образов, возможно создание широкодиапазонного помехоустойчивого аудиокодека, работающего на скоростях от *1* до *64 Кбод* с плавной адаптацией к пропускной способности канала связи. С учетом использования голосовой базы данных конкретного диктора можно достичь нижней границы в *400 – 600 бит/с*, приближаясь вплотную к теоретическому пределу – *70 бит/с*.

С использованием предложенного подхода ОАС можно генерировать разнообразные речеподобные сигналы, аудиомаркеры с заданными характеристиками, что позволит подтверждать подлинность голосовых команд и сообщений, выявлять смонтированные «фейковые» новости, создавать новые языки общения людей и машин, помогать общению больных пациентов с поврежденным горловым и голосовым аппаратом.

Одним из главных применений полутонового и бинарного ОАС является создание на его основе высоконадежных удобных (бесконтактных) конвергенционных технологий удаленной многомодальной (парольная фраза плюс рукописная подпись плюс изображение «лица, портрета» в спектрах речеподобных сигналов) динамической аутентификации пользователей значимых информационных ресурсов, банковских и иных услуг.

Бинарный и полутоновой ОАС может быть использован для решения задач адаптивного компандирования РС в условиях ограничений пропускной способности канала голосовой связи, шумоочистки РС через шумоочистку их графических образов, создания новых нетрадиционных методов цифрового и аналогово-цифрового закрытия РС и других приложений защиты речевой информации.

На основе «бинарного» анализа/синтеза могут быть сформированы новые подходы к проблемам речевых трансформаций, в том числе их частотно-временной и пошаговой модификации, а также к эффективному речевому распознаванию через распознавание бинарных графических образов.

СПИСОК ЛИТЕРАТУРЫ:

1. Устинов, Роман А. Особенности современных систем защиты речевой информации. Безопасность информационных технологий, [S.l.], v. 24, n. 4, p. 71-79, nov. 2017. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/279>. Дата доступа: 24 jan. 24.01.2019. doi:http://dx.doi.org/10.26583/bit.2017.4.08.
2. Дворянкин С.В. Речевая подпись. М.: РИО МТУСИ. 2003. 184 с.
3. Дворянкин С.В., Нагорных И.М. К вопросу о технологии преобразования звук – изображение – звук. // Спецтехника и связь. 2013. № 1. С. 28 – 32.
4. Алюшин В.М., Дворянкин С.В. Технологии образного анализа в задачах цифровой обработки речевой информации. // Научная визуализация. 2013. Т. 5. № 3. С. 75 – 88.
5. Селянкин В.В., Скороход С.В. Анализ и обработка изображений в задачах компьютерного зрения: // Учебное пособие. Таганрог: Издательство Южного федерального университета. 2015. 84 с.
6. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2012. 1104 с.
7. McAulay R.J., Quatieri T.F, Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE Trans. on Acoust., Speech and Signal Processing. – 1988. Vol. l. ASSP-34. P. 744 – 754.
8. Лихачев Д.С., Петровский А.А. Низкоскоростной кодер на основе модели слуха человека. // ЗАО АВТЭКС Санкт-Петербург. 5-я Международная конференция «Цифровая обработка сигналов и ее применение», DSPA-2003.
9. Ghitza O., Auditory Nerve Representation as a Basis for Speech Processing, Advances in Speech Signal Processing, edited by Sadaki Furui, Tokyo, Japan. P. 453 – 485.

10. Moore B.C.J. An Introduction to the Psychology of Hearing, Sixth Edition. // B.C.J. Moore. Leiden: Boston Brill. 2012. 441 p.
11. Устинов Р.А., Алюшин А.М., Дворянкин Н.С. Особенности формирования бинарных изображений аудиомаркеров при организации защищенного документооборота кредитно-финансовых организаций // XVII Всероссийская конференция «Технологии информационной безопасности в деятельности органов внутренних дел». Сборник докладов М.: Московский университет МВД России имени В.Я. Кикотя, 2018. С. 17 – 24.
12. Устинов Р.А. Управление пропускной способностью голосового канала связи через обработку и бинаризацию изображений динамических сонограмм // Экономика: вчера, сегодня, завтра. 2018 Т. 8 № 8А. С. 426 – 436.
13. Алюшин, Александр М.; Дворянкин, Сергей В. Использование речевых технологий для защиты документооборота. Безопасность информационных технологий, [S.l.], v. 24, n. 2, p. 6-15, june 2017. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/100/294>. Дата доступа: 25 jan. 2019. doi:http://dx.doi.org/10.26583/bit.2017.2.01.
14. Дворянкин, Никита С. Анализ методов скрытного маркирования голосовых команд дистанционного речевого управления для подтверждения их подлинности. Безопасность информационных технологий, [S.l.], v. 24, n. 1, p. 18-27, apr. 2017. ISSN 2074-7136. Доступно на: <https://bit.mephi.ru/index.php/bit/article/view/51>. Дата доступа: 24 jan. 2019. doi:http://dx.doi.org/10.26583/bit.2017.1.03.
15. Устинов Р.А. Применение методов обработки и бинаризации изображений спектрограмм речевого сигнала в задачах речевой стеганографии // Инновационные подходы в современной науке. Сборник статей по материалам XXXI международной научно-практической конференции. 2018. С. 75 – 80.
16. Wu Zhijun Information Hiding in Speech Signals for Secure Communication. Science Press. 2015. 183 p.

REFERENCES:

[1] Ustinov R.A. Specific features of modern voice protection systems. IT Security (Russia), [S.l.], v. 24, n. 4, p. 71-79, nov. 2017. ISSN 2074-7136. Available on: <https://bit.mephi.ru/index.php/bit/article/view/279>. Access date: 24 jan. 2019. doi:http://dx.doi.org/10.26583/bit.2017.4.08. (in Russian)
[2] Dvoryankin S.V. Speech signature. M. RIO. 2003. 184 p. (in Russian)
[3] Dvoryankin S.V., Nagornyh I.M. On the issue of technology of conversion of sound - image – sound. Spectekhnika i svyaz'. 2013. № 1. P. 28 – 32. (in Russian)
[4] Alyushin V.M., Dvoryankin S.V. Technologies of imaginative analysis in the tasks of digital processing of speech information. Nauchnaya vizualizaciya. 2013. T. 5. № 3. P. 75 – 88. (in Russian)
[5] Selyankin V.V., Skorohod S.V. Analysis and image processing in computer vision tasks: Tutorial. Taganrog: Izdatelstvo Yuzhnogo federalnogo universiteta. 2015. 84 p. (in Russian)
[6] R.Gonzalez, R. Woods. Digital Image Processing. M.: Tekhnosfera, 2012. 1104 p. (in Russian)
[7] McAulay R.J., Quatieri T.F, Speech Analysis/Synthesis Based on a Sinusoidal Representation, IEEE Trans. on Acoust., Speech and Signal Processing. – 1988. Vol. ASSP-34. P. 744 – 754.
[8] Lihachev D.S., Petrovskij A.A. Low-speed encoder based on human hearing. ZAO AVTEHKS Sankt-Peterburg. 5-ya Mezhdunarodnaya konferenciya «Cifrovaya obrabotka signalov i ee primenenie», DSPA-2003. (in Russian)
[9] Ghitza O., Auditory Nerve Representation as a Basis for Speech Processing, Advances in Speech Signal Processing, edited by Sadaki Furui, Tokyo, Japan. P. 453 – 485.
[10] Moore B.C.J. An Introduction to the Psychology of Hearing, Sixth Edition. B.C.J. Moore. - Leiden: Boston Brill. - 2012. 441 p.
[11] Ustinov R.A., Alyushin A.M., Dvoryankin N.S. Features of the formation of binary images of audio markers in the organization of protected document management of credit and financial organizations. XVII «Vserossijskaya konferenciya «Tekhnologii informacionnoj bezopasnosti v deyatelnosti organov vnutrennih del». Collection of reports – M.: Moskovskij universitet MVD Rossii imeni V.Y. Kikotya, 2018. P. 17 – 24. (in Russian)
[12] Ustinov R.A. (2018) Upravleniye propusknoy sposobnostyu golosovogo kanala svyazi cherez obrabotku i binarizatsiyu izobrazheniy dinamicheskikh sonogramm [Management of the ventilation ability of the voice communication channel through the processing and binarization of images of dynamic sonograms]. Ekonomika: vchera, segodnya, zavtra [Economics: Yesterday, Today and Tomorrow], 8 (8A). P. 426 – 436. (in Russian)
[13] Alyushin A.M, Dvoryankin S.V. The Use of Speech Technology to Protect the Document Turnover. IT Security (Russia), [S.l.], v. 24, n. 2, p. 6-15, june 2017. ISSN 2074-7136. Available on: <https://bit.mephi.ru/index.php/bit/article/view/100/294>. Access date: 25 jan. 2019. doi:http://dx.doi.org/10.26583/bit.2017.2.01. (in Russian)
[14] Dvoryankin N.S. Analysis methods of secretive labelling voice commands for remote voice control to confirm their authenticity. IT Security (Russia), [S.l.], v. 24, n. 1, p. 18-27, apr. 2017. ISSN 2074-7136. Available on: <https://bit.mephi.ru/index.php/bit/article/view/51>. Access date: 24 jan. 2019. doi:http://dx.doi.org/10.26583/bit.2017.1.03. (in Russian)
[15] Ustinov R.A. Application of processing methods and binarization of images of speech signal spectrograms in speech steganography tasks. Innovative approaches in the modern science. Proceedings of XXXI international scientific-practical conference. 2018. P. 75 – 80. (in Russian)
[16] Wu Zhijun Information Hiding in Speech Signals for Secure Communication. Science Press. 2015. 183 p.

Поступила в редакцию – 13 декабря 2019 г. Окончательный вариант – 28 февраля 2019 г.
Received – December 13, 2019. The final version – February 28, 2019.