

УДК: 50.37.23

ИНФОРМАТИВНЫЙ ПРИЗНАК ПЕРЕДАТОЧНЫХ ФУНКЦИЙ МОДЕЛЕЙ РЕЧЕОБРАЗОВАНИЯ ПРИ ИСПОЛЬЗОВАНИИ НИЗКОСКОРОСТНЫХ КОДЕРОВ РЕЧИ

Канд. техн. наук. Д.Н. Чубатый, канд. техн. наук Д.А. Клецков, В.В. Кузьмин

Ph.D. D.N. Chubaty, Ph.D. D.A. Kletskov, V.V. Kuzmin

Военный университет радиоэлектроники (г. Череповец)

В настоящее время в современном обществе в области информационной безопасности у злоумышленников прослеживается тенденция к несанкционированному доступу к персональным данным личности человека с целью передачи информации третьим лицам, таргетирование рекламным материалом, проведения статистических исследований на основе добытых данных. Статья посвящена вопросам, затрагивающим интересы систем идентификации, которые обусловлены применением широкого круга практических задач, условно представленных тремя классами: системами военного или другого специального назначения, коммерческие системы и демонстрационные системы. Материалы, опубликованные в данной статье, будут интересны специалистам, занимающимся исследованиями речеобразования при использовании низкоскоростных кодеров речи, а именно голосовой биометрической идентификации.

Ключевые слова: биометрическая идентификация, низкоскоростные кодеры речи, несанкционированный доступ, декодирование сигнала, персональные данные.

Currently, in modern society in the field of information security, attackers have a tendency to unauthorized access to personal personal data of a person in order to transfer information to third parties, target advertising materials, conduct statistical research based on obtained data. The article deals with issues affecting the interests of identification systems, which are due to the application of a wide range of practical tasks, conditionally presented by three classes: military or other special purpose systems, commercial systems and demonstration systems. The materials published in this article will be of interest to those skilled in speech research using low-speed speech encoders, namely voice biometric identification.

Keywords: biometric identification, low-speed speech encoders, unauthorized access, signal decoding, individual data.

Основным преимуществом применения голосовой биометрической идентификации перед биометрическими системами на основе анализа папиллярных узоров или сетчатки глаза состоит в возможности получения и передачи биометрических данных без применения специализированной и дорогостоящей аппаратуры съема биометрической информации: достаточно иметь

телефон или микрофон, подключенный к компьютеру. По состоянию на декабрь 2019 года, по данным компании Cisco, в мире насчитывается более 4 миллиардов пользователей мобильной связи. Поэтому одним из перспективных направлений реализации технологий идентификации по голосу является предоставление удаленного доступа к различным ресурсам с использовани-

ем телефонных каналов сотовых систем связи. Речевые системы не являются самостоятельными, они встроены в некую «большую» систему, выполняющую конкретную практическую задачу, например: проверка прав доступа к различным системам (базам данным, каналам связи, помещениям, устройствам и механизмам, банковским счетам и т.д.), криминалистическая экспертиза [1].

Голосовая биометрическая идентификация является достаточно эффективным средством предотвращения несанкционированного доступа к данным и позволяет предотвратить такие угрозы информационной безопасности, как: «хищение личности» — использование злоумышленником персональных данных человека, полученных из клиентских баз данных с целью приобретения товаров и услуг, получения кредитов и совершения других неправомерных действий; хищение ценной корпоративной информации сотрудниками компании или внешними злоумышленниками посредством несанкционированного доступа к корпоративным данным.

Большинство существующих и перспективных стандартов мобильной связи при обработке речевых сигналов используют низкоскоростное кодирование, которое позволяет обеспечить увеличение общей пропускной способности системы связи и снизить стоимость предоставления одного канала. В данных случаях это приводит также и к снижению качества воспроизводимой речи, следовательно, и к усложнению задачи идентификации абонента по голосу. Существующие методы идентификации абонента по голосу характеризуются наличием обязательного этапа декодирования речевых сигналов, что требует больших временных и операционных затрат на обработку. Это обуславливает необходимость поиска новых информативных признаков, а также разработки метода идентификации абонентов в сотовых системах связи на объектах мониторинга без предварительного декодирования речевого сигнала. В настоящее время в сотовых системах связи наибольшее широкое распространение получили гибридные кодеры. Формирование речевых сигналов с использованием этих кодеров осуществляется на основе линейного предсказания, а повсеместное их использование обусловлено тем, что позволяет существенно снизить скорость передачи при

приемлемом качестве синтеза сигнала. Линейное предсказание является одним из наиболее эффективных методов анализа речевых сигналов. Этот метод считается доминирующим при оценке параметров речи, таких как период и частота основного тона, форманты, функция площади речевого тракта, используемых при идентификации по голосу, а также при сокращенном представлении речи с целью ее низкоскоростной передачи [2, 3, 4]. Основные положения данного метода хорошо согласуются с моделью формирования речевых сигналов [2, 4, 8], структурная схема которой представлена в виде модели речевого сигнала на основе линейного предсказания, наиболее удобной для решения задач линейного предсказания (далее — модель) (рис. 1). Эта система возбуждается импульсной последовательностью для вокализованных звуков и шумом для невокализованных.

Данная модель имеет следующие параметры: классификатор вокализованных и невокализованных звуков, период основного тона для вокализованных сегментов, коэффициент усиления G и коэффициенты a_i цифрового фильтра. Все эти параметры во времени изменяются медленно. Главное достоинство этой модели заключается в том, что параметры G и a_i можно оценить с использованием очень эффективных с вычислительной точки зрения алгоритмов [2, 9]. Для модели отсчет речевого сигнала S_n связан с сигналом возбуждения u_n разностным уравнением вида:

$$S_n = \sum_{i=1}^{M_{bn}} a_i S_{n-i} + Gu_n. \quad (1)$$

Линейный предсказатель с коэффициентами a_i определяется как система, на выходе которой имеем:

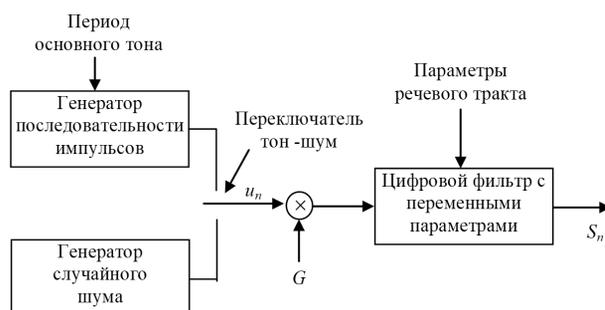


Рис. 1. Модель речевого сигнала на основе линейного предсказания

$$S'_n = \sum_{i=1}^{M_{kn}} a_i S_{n-i}, \quad (2)$$

где a_i — коэффициент взвешивания предшествующих выборок, или коэффициент линейного предсказания; M_{kn} — число коэффициентов предсказания, или порядок предсказания. Чем быстрее убывают корреляционные связи, тем меньшую роль играют ранние отсчеты сигнала, т.е. отстоящие от n -го отсчета на большее расстояние. Обычно, таких отсчетов используется 8...12 ($M_{kn}=8...12$) [2, 4]. Разность между действительным S_n и предсказанным S'_n значениями речевого сигнала определяет ошибку предсказания e_n , или остаток предсказания, и определяется выражением:

$$e_n = S_n - S'_n = S_n - \sum_{i=1}^{M_{kn}} a_i S_{n-i}. \quad (3)$$

В результате z — преобразования этого разностного уравнения получается:

$$E(z) = S(z) - S'(z) = S(z) \cdot A(z). \quad (4)$$

где $A(z)$ — преобразование этого разностного уравнения имеет вид:

$$A(z) = 1 - \sum_{i=1}^{M_{kn}} a_i z^{-i}, \quad (5)$$

где z^{-i} — соответствует разложению речевого сигнала при задержке на один отсчет.

Уравнение (4) представляет собой математическую модель ошибки линейного предсказания речи в области z — изображений [2, 3, 10]. Если речевой сигнал $S(z)$ подается на вход обратного фильтра с характеристикой $A(z)$, коэффициенты которого определяются на основе анализа речевого сигнала, то выходным сигналом будет функция $E(z)$, которая используется для возбуждения синтезатора речевого сигнала. Таким образом, $E(z)$ может интерпретироваться как выходной сигнал фильтра, передаточная функция которого описывается выражением:

Частотная характеристика фильтра анализатора $A(z)$ может быть величиной обратной частотной характеристике фильтра синтезатора речевого сигнала (голосового тракта), поэтому фильтр $A(z)$ называется инверсным. Следовательно, при подаче на вход фильтра $A(z)$ речевого сигнала, при оптимальном выборе его па-

раметров, сигнал на выходе будет представлять собой сигнал возбуждения $E(z)$, подобный сигналу возбуждения на входе фильтра голосового тракта. Инверсная фильтрация речевого сигнала создает декоррелированный сигнал остатка, и чем выше при этом порядок фильтра, тем меньше остаточная корреляция и спектр сигнала остатка в большей степени приближается к спектру белого шума. Из уравнения (5) получается модель синтезирующего фильтра:

$$S(z) = E(z) / A(z) = E(z) \cdot H(z), \quad (6)$$

где $H(z)$ — передаточная характеристика приемного устройства (синтезирующего фильтра):

$$H(z) = G / A(z) = G / \left(1 - \sum_{i=1}^{M_{kn}} a_i z^{-i} \right). \quad (7)$$

Основная задача анализа на основе линейного предсказания заключается в непосредственном определении параметров a_i по речевому сигналу с целью получения оценок его спектральных свойств, путем использования уравнения (7). Полученные параметры составляют основу для описания сигнала. Коэффициент усиления G и коэффициенты линейного предсказания рассчитываются на кадре длительностью около 10–30 мс. На таком сравнительно коротком интервале речевой сигнал может считаться стационарным [2, 11, 12].

Рассмотрим структуру кадров для кодера RPE-LTP (Regular-Pulse Excited LPC with Long-Term Predictor — линейное предсказание с регулярным импульсным возбуждением и долгосрочным предсказателем), реализуемого стандартом GSM 6.10 [6]. В структуре кадра можно выделить несколько групп коэффициентов. Так, визуально кадр RPE-LTP состоит из начальной части кадра длиной 36 бит и 4-х подкадров длиной по 56 бит. Каждый подкадр в свою очередь состоит из 2-х блоков по 17 и 39 бит. Визуальное изображение сигнала, оригинальное название и порядок расположения коэффициентов приведены на рис. 2 и 3 соответственно.

Проведенный анализ особенностей формирования речевых сигналов, преобразованных низкоскоростными кодерами, на примере кодера RPE-LTP — показал, что для передачи по каналу связи используются параметры — коэффициенты логарифмического отражения площадей.



Рис. 2. Визуально-графическое представление речевого сигнала (кодер RPE-LTP), кадр 260 бит

Параметры, относящиеся ко всему кадру								1-ый подкадр					2-ой подкадр		3-ий подкадр		4-ый подкадр	
1-6	7-12	13-17	18-22	23-26	27-30	31-33	34-36	37-43	44-45	46-47	48-53	54-56	... 92	93...148	149...204	205...260		
g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	L_1	A_1	V_1	G_1	x_1	... x_{13}	$L_{2...} x_{13}$	$L_{3...} x_{13}$	$L_{4...} x_{13}$		

Рис. 3. Распределение параметров речи по битам кадра кодера RPE-LTP:

$g_1...g_8$ — коэффициенты логарифмического отражения площадей (LAR); $L_1...L_4$ — задержка;
 $A_1...A_4$ — коэффициент предсказания; $V_1...V_4$ — номер последовательности;
 $G_1...G_4$ — максимальная амплитуда; $x_1...x_{13}$ — нормированные амплитуды импульсов

Эти параметры рассчитываются исходя из первоначально оцененных коэффициентов линейного предсказания [2, 3, 6]. Таким образом, для гибридных кодеров, используемых в сотовых системах связи, характерно наличие группы параметров речи, в основе получения которых лежит схема линейного предсказания.

С другой стороны, в большинстве существующих систем идентификации абонента по голосу, для выделения информативных признаков используется кратковременный анализ. Сигнал разбивается на временные окна фиксированного размера, на которых, как предполагается, параметры сигнала не меняются. Для более точного представления сигнала между окнами делают перекрытие, равное половине длины окна. Затем к каждому окну применяются алгоритмы извлечения признаков, в виде коэффициентов кепстра, либо по передаточной функции речевого тракта, найденной на основе модели линейного предсказания [2, 3, 7]. Используют следующие параметры, полученные на основе метода линейного предсказания [2, 7]: коэффициенты логарифмического отношения площадей (LAR — log area ratios); перцептивные коэффициенты логарифмического отношения площадей (PLAR — perceptive log area ratios); мел-частотные кепстральные коэффициенты (MFCC — mel-frequency cepstral coefficients); коэффициенты линейного предсказания (LPC — linear prediction coefficients); линейные предсказывающие кепстральные коэффициенты

(LPCC — linear prediction cepstral coefficients); изменения линейных предсказывающих кепстральных коэффициентов (Δ LPCC — delta linear prediction cepstral coefficients); нормированные линейные предсказывающие кепстральные коэффициенты (NLPCCC — normalized linear prediction cepstral coefficients); линейно-частотные кепстральные коэффициенты (LFCC — linear-frequency cepstral coefficients); линейные спектральные пары или частоты (LSP — linear spectral pair или LSF — linear spectral frequency). Важным достоинством модели линейного предсказания является то, что она может быть легко преобразована в модель речеобразования, полагаясь при этом на физиологические принципы, представляющие основу исследований во многих областях, имеющих отношение к речи [4]. Одна из наиболее распространенных моделей речеобразования основана на предположении, что голосовой тракт можно представить в виде неоднородной акустической трубы, которая простирается от голосовой щели до губ и изменяется по форме с течением времени. Основными анатомическими компонентами, вызывающими это изменение, являются губы, челюсти, язык и нёбная занавеска (рис. 4).

Модель речеобразования, включающая в себя последовательное соединение N_c — секций акустических труб без потерь, где A_i — «функция площади», т.е. площадь поперечного сечения в направлении, перпендикулярном продольной оси трубы (рис. 5). Передаточная функ-

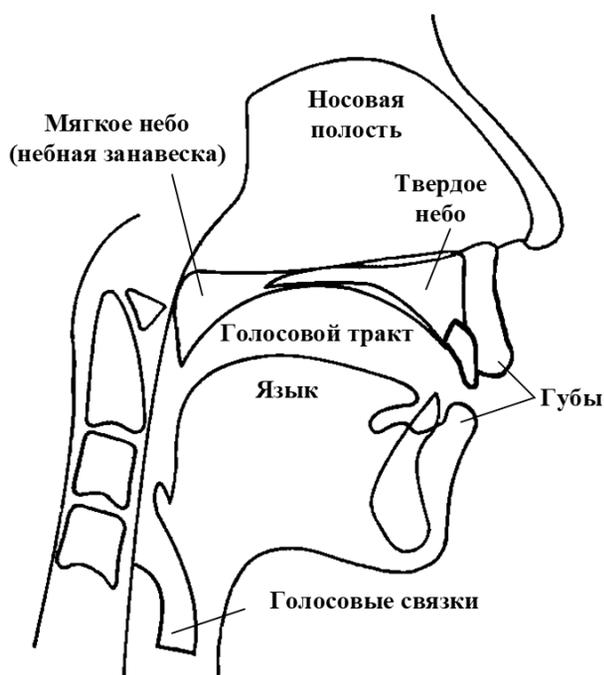


Рис. 4. Схематичное изображение речевого аппарата человека

ция модели с трубами без потерь имеет столько элементов задержки, сколько секций в модели. Передаточная функция имеет полюса и не имеет нулей. Эти полюса определяют резонансы или форманты модели [2]. Пусть число коэффициентов линейного предсказания — M_{kn} в модели речевого сигнала равно числу секций — N_c в модели речеобразования в виде акустической трубы без потерь. Тогда, передаточная функция синтезирующего фильтра приемного устройства при низкоскоростном кодировании речи и пере-

даточная функция речевого тракта ИРС, в виде акустической трубы без потерь, равны.

Учитывая также, что коэффициенты отражения и коэффициенты частной корреляции связаны соотношением [2]:

$$r_i = -k_i.$$

Тогда, передаточная функция синтезирующего фильтра приемного устройства при низкоскоростном кодировании речи и передаточная функция речевого тракта ИРС, в виде акустической трубы без потерь, равны. Коэффициенты отражения r_i на рис. 5 связаны с площадями поперечного сечения соотношением [2]:

$$r_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i}, \quad (9)$$

где A_i — «функция площади», т.е. площадь поперечного сечения в направлении, перпендикулярном продольной оси трубы. Передаточная функция такой системы в предположении, что коэффициент отражения от источника возбуждения $r_G = 1$, т.е. сопротивление источника предполагается бесконечно большим, имеет вид [2]:

$$V(z) = G / D(z), \quad (8)$$

где $D(z)$ — удовлетворяет соотношениям:

$$D_0(z) = 1; \quad (9)$$

$$D_i(z) = D_{i-1}(z) + r_i z^{-i} D_{i-1}(z^{-1}); \quad (10)$$

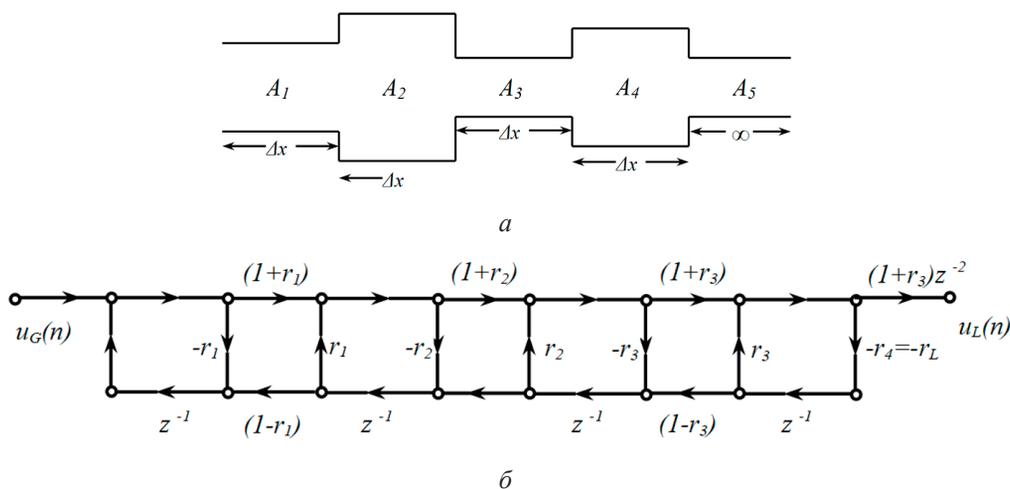


Рис. 5. а — модель трубы без потерь, бесконечной длины; б — граф прохождения сигнала при бесконечном сопротивлении источника

$$D_i(z) = D_{N_c}(z); \quad (11)$$

$$D_i(z) = 1 - \sum_{i=1}^{N_c} a_i z^{-i}. \quad (12)$$

Передаточная функция синтезирующего фильтра, используемая при кодировании речевых сигналов:

$$H(z) = G / A(z),$$

где $A(z)$ может быть получена на основе линейного предсказания с использованием рекурсивной процедуры [2]:

$$A_0(z) = 1;$$

$$A_i(z) = A_{i-1}(z) - k_i z^{-i} A_{i-1}(z^{-1});$$

$$A(z) = A_M(z),$$

где параметры k_i — коэффициенты частной корреляции. Из выражений (1–7) и (8–12) видно, что передаточная функция:

$$H(z) = G / A(z) = G / \left(1 - \sum_{i=1}^{M_{kn}} a_i z^{-i} \right),$$

полученная на основе линейного предсказания, имеет тот же вид, что и передаточная функция акустической трубы без потерь, имеющей N_c секций

$$V(z) = G / D(z) = G / \left(1 - \sum_{i=1}^{N_c} a_i z^{-i} \right),$$

а коэффициенты отражения r_i и коэффициенты частной корреляции k_i зависят от коэффициентов линейного предсказания a_i . Таким образом, если число коэффициентов линейного предсказания M_{kn} равно числу секций N_c в модели речеобразования в виде акустической трубы и учитывая, что $r_i = -k_i$ [5], то передаточная функция речевого тракта в виде акустической трубы без потерь и передаточная функция синтезирующего фильтра приемного устройства при низкоскоростном кодировании речи равны:

$$H(z) = V(z).$$

Следовательно, параметры речи, используемые при низкоскоростном кодировании и по-

лучаемые на основе метода линейного предсказания: коэффициенты линейного предсказания, коэффициенты логарифмического отношения площадей, линейные спектральные частоты и др. зависят от коэффициентов отражения r_i , характеризующих особенности речевого тракта и могут быть использованы при решении задачи идентификации абонента в сотовых системах связи на объектах мониторинга [2, 7, 8, 11]. Таким образом, можно сделать вывод, что применение данных параметров позволит проводить идентификацию абонентов до этапа декодирования и значительно сократить вычислительные ресурсы при заданном качестве распознавания.

Литература

1. Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. 2012. Т. 12. № 1. С. 1–30.
2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов / Под ред. М.В. Назарова и Ю.Н. Прохорова. — М.: Радио и связь. 1981. 496 с.
3. Маркел Дж.Д., Грэй А.Х. Линейное предсказание речи: Пер. с англ / Под ред. Прохорова Ю.Н. и Звездина В.С. — М.: Связь. 1980. 308 с.
4. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи / Под ред. Шелухина О.И. — М.: Радио и связь. 2000. 456 с.
5. Ратынский М.В. Основы сотовой связи. / Под ред. Д.Б. Зимина. 2-е изд. — М.: Радио и связь. 2000. 248 с.
6. Первушин Е.А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование: Сборник научных трудов. Вып. 24 / Под ред. А.К. Гуца. — Омск: ОмГУ. 2011. С. 41–54.
7. Кузьмин В.В., Козлов С.Ю. Методика построения маршрутов движения объектов в районе сбора информации // Известия Тульского государственного университета. Технические науки. 2020. № 9. С. 225–231.
8. Кузьмин В.В. Структурно-функциональная модель объекта радиомониторинга на основе комплексирования информации с применением математического аппарата теории информационного поиска // Вопросы оборонной техники. Серия 16. 2018. № 9–10 (123–124). С. 138–144.

9. Еремеев И.Ю., Клецков Д.А., Шишкалов А.В. Информативный признак для распознавания служебных пакетов канала управления спутниковых систем связи с многостанционным доступом и временным разделением абонентов // *Информация и космос*. 2015. № 3. С. 50–61.

10. Егоров А.С., Клецков Д.А. Применение методов технического маскирования в пиринговых файлообменных сетях // *Инновационная деятельность в вооруженных силах Российской Федерации. Труды всеармейской научно-практической конференции. Министерство обороны Российской Федерации, Международная академия авторов научных открытий и изобретений (Санкт-Петербургское отделение). Военная академия связи*. 2019. С. 116–119.

11. Маковский В.Н., Чеботарь И.В., Кузьмин В.В., Гайчук Ю.Н. Методы оценивания пространственных характеристик мобильных абонентов // *Антенны*. 2015. № 11. С. 38–42.

12. Гудков А.А., Клецков Д.А., Кузьмин В.В., Удальцов Н.П. Модель распознавания объектов радиомониторинга в иерархических системах управления // *Известия Тульского государственного университета. Технические науки*. 2018. № 1. С. 283–291.

References

1. Sorokin V.N., V'yugin V.V., Tananakin A.A. Personal identification by voice: an analytical review // *Information processes*. 2012. Vol. 12. № 1. P. 1–30.

2. Rabiner L.R., Shafer R.V. Digital processing of speech signals // Edited by M.V. Nazarov and Yu.N. Prokhorov. — M.: Radio and communications. 1981. 496 p.

3. Markel J.D., Gray A.H. Linear speech prediction // *TRANS. from English*. Edited by Yu.N. Prokhorov and V.S. Zvezdin. — Moscow: Svyaz. 1980. 308 p.

4. Shelukhin O.I., Lukyantsev N.F. / Ed. Shelukhina O.I. Digital speech processing and transmission. — M.: Radio and communications. 2000. 456 p.

5. Ratynsky M.V. Fundamentals of cellular communication // Under the editorship of D.B. Zimin. Izd. № 2. — M.: Radio and communications. 2000. 248 p.

6. Pervushin E.A. Overview of the main methods of speaker recognition // *Mathematical structures and modeling: a Collection of scientific papers*. № 24 / Ed. by A.K. Gutsa. — Omsk: OmSU. 2011. P. 41–54.

7. Kuzmin V.V., Kozlov S.U. Methodology for building routes for objects in the area of information collection // *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskiye nauki*. 2020. № 9. P. 225–231.

8. Kuzmin V.V. Structural and functional model of the monitoring object based on information aggeration using the mathematical apparatus of information search theory // *Questions of engineering. Series 16*. 2018. № 9–10 (123–124). P. 138–144.

9. Eremeev I.U., Kletskov D.A., Schischkalov A.V. Informative feature for recognizing service packets of the control channel of satellite communication systems with multi-station access and time division of subscribers // *Informazija i kosmos*. 2015. № 3. P. 50–61.

10. Egorov A.S., Kletskov D.A. Application of technical masking methods in peer-to-peer file-sharing networks // *Innovazionnaja dejatel'nost' v Voorugennih Silah Rossiiskoi Federacii. Trudi vsearmeiskoi nauchnoprakticheskoi konferencii. Ministerstvo obroni Rossiiskoi Federacii, Megdunarodnaja akademija avtorov nauchnih otkritii i izobretenii (Sankt-Peterburgskoe otdelenie). Voennaja akademija svjazi*. 2019. P. 116–119.

11. Makovskiy V.N., Chebotar I.V., Kuzmin V.V., Gaychuk Yu.N. Methods for estimating the spatial characteristics of mobile subscribers // *Antennas*. 2015. № 11. P. 38–42.

12. Gudkov A.A., Kletskov D.A., Kuzmin V.V., Udaltsov N.P. Model of recognition of radiomonitoring objects in hierarchical control systems // *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskiye nauki*. 2018. № 1. P. 283–291.