УДК 004.056 Вестник УрФО № 1(27) / 2018, с. 34–41

Дресвянин П. Д., Сафиуллин Н. Т., Поршнев С. В.

О ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА ЭМПИРИЧЕСКОЙ МОДОВОЙ ДЕКОМПОЗИЦИИ ДЛЯ ИДЕНТИФИКАЦИИ АВТОРА РЕЧЕВОГО СИГНАЛА

В статье проводится анализ проблем, связанных с идентификацией источников акустических сигналов в случае их использовании в качестве паролей при авторизации пользователей интеллектуальных информационных систем (ИС). Приведен обзор актуальных решений, основанных на использовании биометрических показателей, а также обоснован выбор направления их дальнейшего развития в России. Особое внимание уделено необходимости обеспечения информационной безопасности (ИБ) данных решений в связи с активным развитием методов атак на компьютерные технологии, использующие методологии машинного обучения и искусственных нейронных сетей. Предложена новая методика идентификации речевых сигналов, основанная на использовании эмпирической модовой декомпозиции (метода преобразования Хуанга-Гильберта), в которой для идентификации автора цифрового звукового сигнала используется коэффициент линейной корреляции между модами, выделенными в результате его декомпозиции. Продемонстрировано, что предложенная методика достаточно устойчива к шумам, присутствующим в речевом сигнале, и вариациям его длительности.

Ключевые слова: *эмпирическая модовая декомпозиция, идентификация, биометрическая авторизация, цифровая обработка сигналов, голосовой пароль.*

ON THE POSSIBILITY TO USE EMPIRICAL MODE DECOMPOSITION TECHNIQUE FOR SPEECH IDENTIFICATION

The article analyses the issues of speech identification – the process that underlies the biometric voice-password authorization in intellectual informational systems. The actual methods of biometric authorization, based on machine learning and artificial neural networks, are presented; the direction of their development and their necessity for Russia are specified in the paper. Special accent made on the necessity of providing informational security for such tasks due to the development of methods for attacks on identification procedures, specifically based on machine learning and artificial neural network techniques. Because of the analysis, the authors provide an alternative technique for speech identification in biometric authorization intellectual systems by using empirical mode decomposition. The paper shows the possibility of using the linear correlation coefficient between identical (by number) intrinsic modes, received by decomposition, as a measure for one-valued recognition, required for authorization. The authors demonstrate that the proposed technique is quite stable and robust in case of noise and other perturbations of digital speech signal. The complexity and accuracy of this technique can be increased by using time-and-frequency analysis of the received intrinsic modes – an issue, which is scheduled for a future research.

Keywords*: empirical mode decomposition, identification, biometric authorization, digital signal processing, voice password.*

Введение

Интеллектуальные интерфейсы, обеспечивающие на основе идентификации речевого сигнала взаимодействие между пользователем и той или иной технической системой, в настоящее время находят применение в различных областях науки и техники. Отметим, что, если на начальном этапе применения подобные системы рассматривались исключительно как дополнительное средство ввода команд голосом, то теперь распознавание речи является неотъемлемой частью методов авторизации пользователей с помощью биометрических показателей [1]. Данная ситуация способствует быстрому развитию информационных технологий (ИТ) в данной области [2], что подтверждается наличием большого числа программных инструментов в сегменте речевых помощников (Google Now, Amazon Alexa, Яндекс Алиса и др.). Также отметим, что, например, в планах департамента финансовых и ИТ Банка России запланировано использование цифрового профиля пользователя, в котором будут интегрированы биометрические данные и образцы голоса, что обеспечит возможность удаленной идентификации пользователей для обеспечения его доступа к таким видам услуг как страховые, пенсионные и нотариальные [3]. При этом понято, что подобная идентификация пользователей цифрового профиля на основе биометрических данных и образца голоса обеспечивает значительное повышение ИБ соответствующих автоматизированных интеллектуальных систем. Отметим, что данные решения в дальнейшем можно будет тиражировать, в том числе, в сферу государственных и муниципальных цифровых услуг, что особенно актуально, принимая во внимание Постановление Правительства РФ от 24.10.2011 № 861 (ред. от 10.02.2018) «О федеральных государственных информационных системах, обеспечивающих предоставление в электронной форме государственных и муниципальных услуг (осуществление функций)».

В области идентификации параметров речевых сигналов приходится решать задачи, отличающиеся друг от друга своей постановкой.

Например, при разработке тех или иных речевых помощников наиболее актуальной оказывается задача точного распознавания слов, их верного семантического истолкования и извлечения информации из ключевых фраз [2]. В тоже время для систем защиты информации на основе биометрических показателей необходимо установить взаимно однозначное соответствие между данным речевым сигналом и конечным пользователем. При этом, как при решении первой, так и второй задачи, необходимо учитывать, что частотно-временные характеристики любого речевого сигнала обладают высокой вариабельностью, следствием которой оказывается его нестационарность, кроме того в речевых сигналах присутствует шум, как правило, с неизвестной функцией распределения. Отмеченные обстоятельства, вообще говоря, ставят по сомнение правомерность использования стандартные статистические или спектральные методы, необходимым условием применения которых является стационарность анализируемого сигнала [4].

В этой связи поиск новых подходов к решению задачи идентификации речевых сигналов – установления однозначного соответствия между сохраненным в системе доступа речевым паролем и акустическим сигналов, зарегистрированным при его повторном произнесении пароля в других условиях и, соответственно, с возможными искажениями, – является актуальной и обладает несомненной практической ценностью для систем ИБ на основе считывания биометрических показателей.

В настоящее время из-за теоретических ограничений, возникающих при использовании известных спектральных и статистических методов для анализа нестационарных временных рядов, к которым относятся и речевые сигналы, продолжается активный поиск альтернативных методов оценивания их частотно-временных характеристик. В области речевых помощников такой альтернативой стали методы машинного обучения и нейронные сети [2; 5], в которых задачу установления однозначного соответствия между речевым сигналом и некоторым семантическим ключом возлагают на автоматизированные средства поиска оптимального решения (например, при обучении нейронных сетей с обратным градиентным распространением ошибки). Получившийся «черный ящик» содержит в себе невидимые для конечного пользователя зависимости и коэффициенты, формирующие для заданного речевого сигнала некоторый ключ, обеспечивающий в дальнейшем его идентификацию. Каждый из методов, базирующихся на машинном обучении, работает по принципу «фразы-пароля» или «фразы-ключа», поскольку в данных методах изначально отсутствуют средства анализа частотно-временных характеристик речевых сигналов. Однако в силу обобщающей способности подобных алгоритмов они оказываются устойчивыми к шуму и изменениям тембра/тона голоса.

С точки зрения требований ИБ к системам опознавания по голосу, методы, основанные на использовании алгоритмов машинного обучения, обладают существенным недостатком – их входные данные можно подделать таким образом, чтобы получить выходной оптимальный результат даже без знания необходимого ключа. Это обусловлено тем, что конечные связи и коэффициенты найденного машинного решения не могут быть найдены непосредственно из исходной информации напрямую, а калибруются в ходе обобщающего обучения, что приводит к невозможности четкого сопоставления исходной информации с конечным результатом [6].

Подобный недостаток был продемонстрирован на примере задачи классификации изображений на основе алгоритмов искусственных нейронных сетей с помощью атак вида «черный ящик» [7]. Здесь задача классификации изображения состояла в отнесении данного изображения с заданным уровнем достоверности к соответствующему классу или установление соответствия между изображением и некоторым термином (например, названием вида животного). В [7] было показано, что существует некоторая пиксельная маска, при наложении которой на это изображение обученная нейронная сеть будет распознавать уже другой класс понятий. Более того, оказывается возможным наложить на любое изображение такую маску, которая обеспечит заранее ожидаемый результат, то есть фальсификацию ключа. При этом с точки зрения человека, что наиболее важно, изображение не претерпевает существенных изменений, но результат, возращенный обученной искусственной нейронной сетью, в данном случае оказывается совершенно неожиданным. Отметим, что подобных примеров «обмана» нейронных сетей в задачах распознавания речевых сигналов пока не опублико-

вано. Однако наличие такой потенциальной возможности заставляет усомниться в возможности применения данной методики в области ИБ.

Альтернативой методам машинного обучения могут служить адаптивные методы анализа цифровой информации, не накладывающие на исходный сигнал никаких ограничений по его статистическим, спектральным и прочим характеристикам. Одним их таких относительно новых методов является Преобразование Хуанга-Гильберта [8], созданное Н. Хуангом в 1998 г., и состоящее из двух этапов. На первом этапе выполняется эмпирическая модовая декомпозиция (ЭМД) исходного сигнала на компоненты, содержащие ключевые параметры исходного сигнала. На втором этапе с помощью преобразования Гильберта или построения Гильбертова спектра определяются частотно-временные характеристики выделенных компонент и сигнала в целом. Далее в нашей статье будет продемонстрировано, что информации, выделенной с помощью метода ЭМД, оказывается достаточно для идентификации речевых сигналов, то есть для установления взаимно-однозначного соответствия между исходным речевым паролем и его последующим произношением в других условиях, в том числе, с искажениями. Расчет частотно-временных характеристик и выделение из них ключевых параметров речевого сигнала не рассматривались, так как являются более широкой задачей, для которой требуются дополнительные исследования.

Основные сведения о методе ЭМД

Напомним, что метод ЭМД является эмпирическим алгоритмом, который не предъявляет к исходному временному ряду (ВР) требований о стационарности его характеристик, но требует, чтобы значения ВР были заданы в узлах равномерной временной сетки. В основе метода ЭМД лежит построение интерполяционных огибающих кривых ВР, проходящих через локальные максимумы/минимумы ряда, с последующим устранением их среднего из анализируемого сигнала до тех пор, пока остаток не будет удовлетворять некоторому условию, после чего он принимается за компоненту. Метод строится итерационно, что позволяет разложить исходный сигнал в аддитивную сумму характеристических компонент, лежащих в разных частотных областях [8]. Блок-схема полного алгоритма, подробно описанного в [12], представлена на рис. 1.

Несмотря на первые успешные результаты, полученные с помощью данного метода при анализе различных цифровых сигналов [8], ЭМД обладала двумя существенными недостатками: низкой точностью разделения

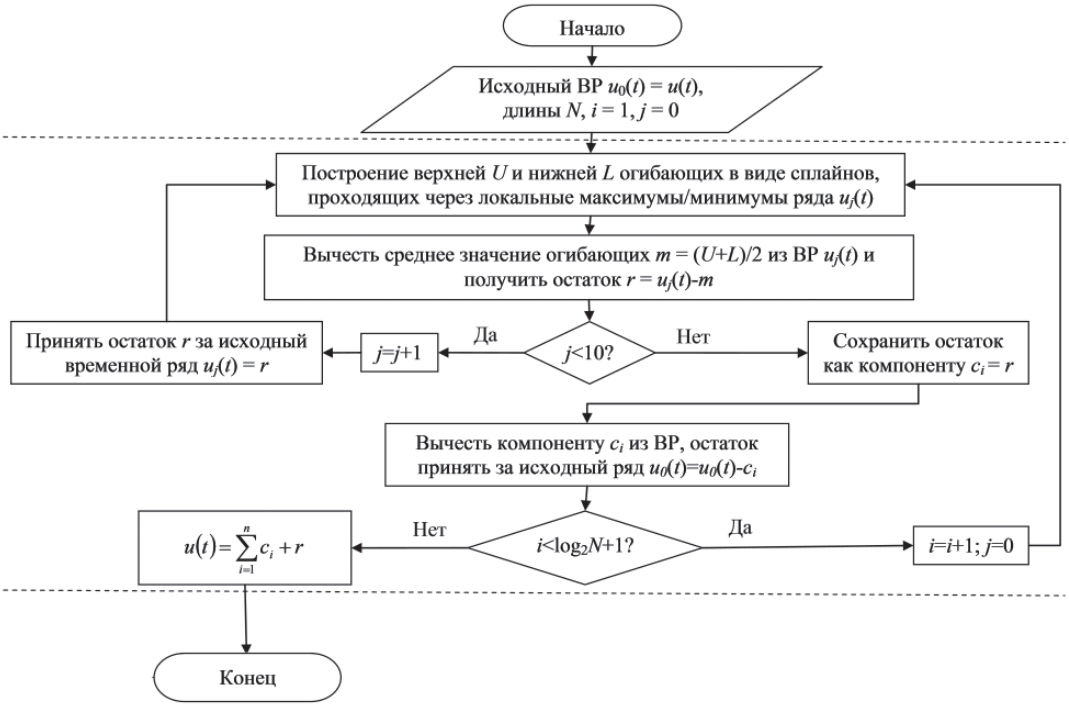

Рис. 1. Блок-схема алгоритма эмпирической модовой декомпозиции (ЭМД)

компонент при наличии в исходном сигнале шума большой мощности [9] и низкой скоростью вычислительного алгоритма в целом [8]. Однако в последние годы в алгоритм ЭМД были внесены существенные изменения. Исходная декомпозиция была доработана до комплементарной ансамблевой эмпирической модовой декомпозиции (CEEMD) [10], которая обеспечила существенное повышение устойчивости метода к шуму. Также удалось достичь значительного сокращения времени вычислений ЭМД, зависящего, однако, от архитектуры его конкретной программной реализации [11], в том числе при использовании технологий параллельных вычислений [12]. Таким образом, сегодня метод ЭМД, в том числе как составная часть Преобразования Хуанга-Гильберта, является инструментом, готовым к использованию для анализа временных рядов и цифровых сигналов.

Методика проведения исследования

Исследование возможности идентификации речевых сигналов на основе метода ЭМД проведено в соответствие с методикой, реализующейся выполнением следующей последовательности действий:

Декомпозиция исходного речевого сигнала *y* длины *T* с частотой дискретизации f_d с помощью метода CEEMD (модификации ЭМД [11; 12]) на характеристические компоненты F_i, общее число которых *N* фиксировано

и известно заранее: $y = \sum_{i=1}^{N} F_i$.

Выбор одной из выделенных компонент анализируемого сигнала для идентификации речевого сигнала. Номер этой компоненты *K* и ее отсчеты временного ряда F_K являются необходимым цифровым ключом $\{K, F_K\}$ для идентификации речевого сигнала. В дальнейшем планируется проверка гипотезы о возможности хранения не самих отсчетов речевого сигнала, а только некоторых его характеристик, по которым можно будет установить однозначное соответствие между ключом и считываемым речевым сигналом.

Выбор некоторого нового речевого сигнала *x*, в качестве претендента на идентификацию в качестве ключа, представляющего собой временной ряд близкой длины, значения которого заданы в узлах временной сетки с такой же частотой дискретизации f_d.

Декомпозиция временного ряда *x* с помощью метода CEEMD на характеристические компоненты G_i, G_i, $x = \sum_{i=1}^{N} G_i$, при этом общее число компонент *N* выбирается таким же, как и у исходного сигнала-пароля *y*.

Сравнение компоненты G_K с номером *K* с соответствующей компонентой-ключом F_K путем вычисления коэффициента корреляции Пирсона. При значении этого коэффициента выше некоторого порогового значения, речевой сигнал *x* принимается за действительный ключ, тем самым подтверждая его идентификацию. В обратном случае речевой образец *x* отвергается, и авторизация не проходит.

Выбор коэффициента корреляции Пирсона для оценивания соответствия компонент F_K и G_K обусловлен его достаточной устойчивостью к временному сдвигу сигналов. Далее будет продемонстрировано, что даже такой простой характеристики оказывается достаточно для подобной задачи с учетом использования алгоритма ЭМД. В дальнейшем планируется усложнение алгоритма с использованием частотно-временных характеристик компонент в качестве ключевых параметров для идентификации соответствующего цифрового ключа.

Анализ экспериментальных результатов

Описанная выше методика исследований была апробирована на стандартном речевом сигнале «speech_dft», из аудио-библиотеки образцов MATLAB Audio System Toolbox. Данный речевой образец содержит запись мужского голоса хорошего качества без существенных помех, длина сигнала 5.1199 секунд, частота дискретизации 22 500 Гц, используется только один аудиоканал (моно-запись).

В аудио-библиотеке MATLAB Audio System Toolbox также имеется готовый искаженный этот же речевой образец худшего качества с высокоуровневым шумом мощности 5 дБ, длина сигнала отличается несущественно (5.0175 секунд), частота дискретизации 22 500 Гц, используется только один аудиоканал (моно-запись). Данные сигналы представлены на рис. 2. Также исходный речевой сигнал для расширенного тестирования предложенного алгоритма искажался вручную следующими способами: наложением шума разной мощности (от 2 до 20 дБ); сокращением общей длины сигнала в меньшую сторону (до 20 %). Коэффициент корреляции между исходным речевым сигналом и его искаженными вариантами не превышает значения в 0.65.

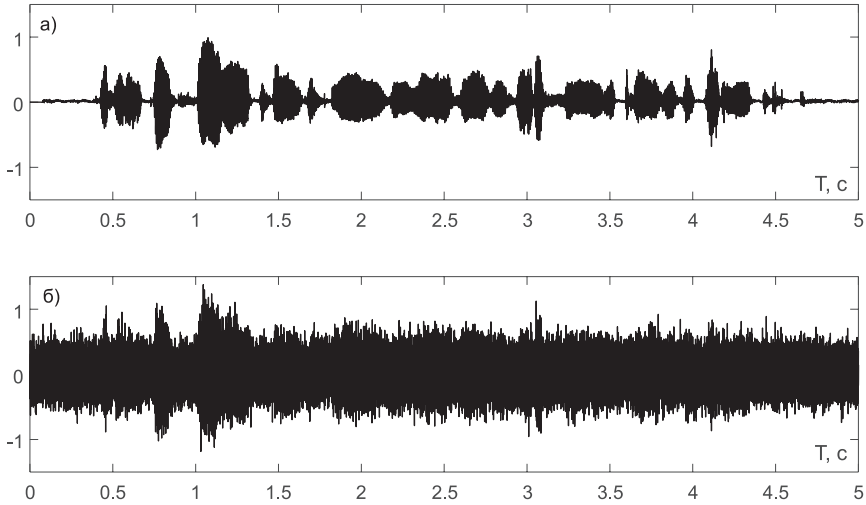

Рис. 2. Исходный речевой сигнал (а) и искаженный сигнал (б)

Результаты пробной идентификации для различных номеров компонент приведены в таблице. Пороговое значение коэффициента корреляции Пирсона для однозначной идентификации было выбрано уровнем выше 0.90.

Средний коэффициент корреляции между характеристической компонентой с номером *K* исходного речевого образца и его искажения

Номер компоненты	Искаженный сигнал (рис. 2б)	Искажение сигнала шумом	Искажение сигнала по длине
1	0.4843	0.2523	0.5002
2	0.4345	0.1291	0.4474
3	0.6970	0.5309	0.7074
4	**0.9760**	**0.9618**	**0.9547**
5	**0.9166**	0.8309	**0.9070**

Из таблицы видно, что в качестве характеристической компоненты для идентификации данного речевого образца лучше всего выбирать компоненту с номером $K = 4$, так как она оказывается наиболее устойчивой ко всем искажениям и обладает наибольшим коэффициентом корреляции между исходным речевым ключом и пробным речевым сигналом.

Компоненты № 4 для исходного речевого сигнала (а) и для искаженного сигнала (б) приведены на рис. 3. Из рис. 3 даже визуально видно, что выделенные компоненты весьма похожи друг на друга.

Также для проверки данного алгоритма проведено сравнение исходного речевого сигнала с другими речевыми рядами (другими фразами) той же длины T с той же частотой дискретизации f_d. Во всех случаях и по всем компонентам коэффициент корреляции не

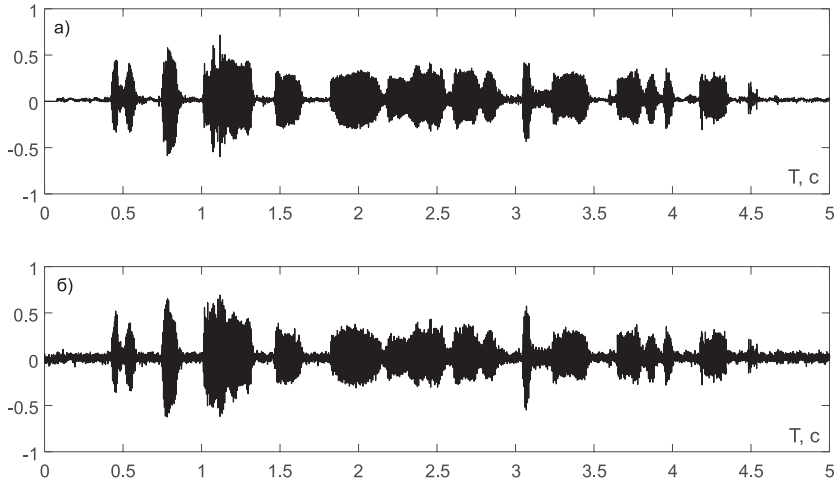

Рис. 3. Компонента под номером $K = 4$ для исходного речевого сигнала (а) и для искаженного сигнала (б)

превысил 0.60, то есть оказывался существенно ниже ожидаемого порогового значения, тем самым обеспечивая некоторую устойчивость к подбору идентификационного ключа.

Таким образом, результаты проведенных экспериментальных исследований предложенной методики подтвердили ее работоспособность. Далее авторы планируют проверить ее устойчивость и точность при использовании различных слов-ключей и различных уровнях акустических помех.

Заключение

Предложена методика идентификации речевых сигналов, использующихся в качестве биометрического пароля доступа к интеллектуальным информационных системам, основанная на методе эмпирической модовой декомпозиции.

Обосновано, что данная методика устойчива к атакам типа «черный ящик» [7].

Приведены экспериментальные результаты, подтверждающие ее работоспособность.

Определены направления дальнейших исследований, цель которых состоит в автоматизации выбора наиболее информативной с точки зрения решаемой задачи компоненты речевого сигнала, а также повышении точности и устойчивости разработанной методики, результаты которых станут предметом последующих публикаций.

Примечания

1. Shari Trewin, etc. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC '12). ACM, New York, NY, USA, – 2012. – pp. 159–168. DOI:10.1145/2420950.2420976

2. Alisa Kongthon, etc. Implementing an online help desk system based on conversational agent. In Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES '09). ACM, New York, NY, USA, – No. 69, – 2009. DOI:10.1145/1643823.1643908

3. Российская газета. Федеральный выпуск №7515 (52). [Электронный ресурс] https://rg.ru/2018/03/13/rossiiane-smogut-brat-kredity-i-otkryvat-vklady-po-vneshnosti-i-golosu.html (Дата обращения 28.03.2018)

4. Сергиенко А. Б. Цифровая обработка сигналов. – 2-е. изд. – СПб. : Питер 2007. – С. 751.

5. Tur, G., De Mori R., Spoken Language Understanding: Systems for Extracting Semantic Information from Speech. – John Wiley & Sons, Ltd – 23 March 2011. – 450 p. DOI: 10.1002/9781119992691

6. M. Hagan, H. Demuth, M. Beale. Neural Network Design. – Amazon Publ. 2nd ed. – 2016. – 1012 p.

7. Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, Ananthram Swami. Practical Black-Box Attacks against Machine Learning. Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE. – 2017. – pp. 506–519.

8. Huang N. E. The Hilbert-Huang transform and its applications / Ed. By S.S. Shen. Interdisciplinary mathematical sciences. 5 Toh Tuck Link, Singapore 596224: World Scientific Publishing Company Co. Pte. Ltd., 2005. – 311 p.

9. Kaslovsky D. N., Meyer F. G. Noise corruption of Empirical Mode Decomposition and its effect on Instantaneous Frequency. Advances in Adaptive Data Analysis. – 2010. – Vol. 2. – No. 3. – P. 373–396.

10. Yeh J.-R., Shieh J.-S., Huang N. E. Complementary Ensemble Empirical Mode Decomposition: A Novel Noise Enhanced Data Analysis Method. Advances in Adaptive Data Analysis. – 2010. – Vol. 2. – No. 2. – P. 135–156.

11. Eftekhar, A., Toumazou, C. & Drakakis, E.M. J Sign Process Syst. – 2013. – Vol. 73. – No. 43. DOI: 10.1007/s11265-012-0726-y

12. Сафиуллин Н. Т. Повышение быстродействия ансамблевой эмпирической модовой декомпозиции распараллеливанием алгоритма // 26-я Международная Крымская конференция «СВЧ-техника и телекоммуникационные технологии». КрыМиКо'2016. – Севастополь, 2016. – С. 593–599.

References

1. Shari Trewin, etc. Biometric authentication on a mobile device: a study of user effort, error and task disruption. In Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC '12). ACM, New York, NY, USA, – 2012. – pp. 159-168. DOI:10.1145/2420950.2420976

2. Alisa Kongthon, etc. Implementing an online help desk system based on conversational agent. In Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES '09). ACM, New York, NY, USA, – No. 69, – 2009. DOI:10.1145/1643823.1643908

3. Russian newspaper. Federal Issue No. 7515 (52). URL: https://rg.ru/2018/03/13/rossiiane-smogut-brat-kredity-i-otkryvat-vklady-po-vneshnosti-i-golosu.html (accessed on 28.03.2018)

4. Sergienko A. B. Digital signal processing. – 2nd. Ed. – 2007. – 751 p.

5. Tur, G., De Mori R., Spoken Language Understanding: Systems for Extracting Semantic Information from Speech. – John Wiley & Sons, Ltd – 23 March 2011. – 450 p. DOI: 10.1002/9781119992691

6. M. Hagan, H. Demuth, M. Beale. Neural Network Design. – Amazon Publ. 2nd ed. – 2016. – 1012 p.

7. Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, Ananthram Swami. Practical Black-Box Attacks against Machine Learning. Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security, Abu Dhabi, UAE. – 2017. – pp. 506-519.

8. Huang N. E. The Hilbert-Huang transform and its applications / Ed. By S.S. Shen. Interdisciplinary mathematical sciences. 5 Toh Tuck Link, Singapore 596224: World Scientific Publishing Company Co. Pte. Ltd., 2005. – 311 p.

9. Kaslovsky D. N., Meyer F. G. Noise corruption of Empirical Mode Decomposition and its effect on Instantaneous Frequency. Advances in Adaptive Data Analysis. – 2010. – Vol. 2. – No. 3. – P. 373–396.

10. Yeh J.-R., Shieh J.-S., Huang N. E. Complementary Ensemble Empirical Mode Decomposition: A Novel Noise Enhanced Data Analysis Method. Advances in Adaptive Data Analysis. – 2010. – Vol. 2. – No. 2. – P. 135–156.

11. Eftekhar, A., Toumazou, C. & Drakakis, E.M. J Sign Process Syst. – 2013. – Vol. 73. – No. 43. DOI: 10.1007/s11265-012-0726-y

12. Safiullin N.T. Povishenie bistrodeystvia ansamblevoi empiricheskoi modovoi decomposicii rasparallelivaniem algoritma // 26th Mezdunarodnaya Crimskaya conference «SVCH-tehnika & telecommunicationnie tehnologii». Crimico 2016. – Sevastopol, 2016. – 593-599 p.

ДРЕСВЯНИН Павел Дмитриевич, студент магистратуры Института Радиоэлектроники и Информационных Технологий Уральского Федерального Университета им. Первого Президента России Б.Н. Ельцина. 620002 г. Екатеринбург, ул. Мира, 32. E-mail: pdresvyanin@bk.ru

САФИУЛЛИН Николай Тахирович, канд. техн. наук, доцент Департамента Информационных Технологий и Автоматики, Института Радиоэлектроники и Информационных Технологий Уральского Федерального Университета им. Первого Президента России Б.Н. Ельцина. 620002 г. Екатеринбург, ул. Мира, 32. E-mail: n.t.safiullin@urfu.ru

ПОРШНЕВ Сергей Владимирович, докт. техн. наук, проф., директор Учебно-научного центра «Информационная безопасность» Института Радиоэлектроники и Информационных Технологий Уральского Федерального Университета им. Первого Президента России Б.Н. Ельцина. 620002 г. Екатеринбург, ул. Мира, 32. E-mail: s.v.porshnev@urfu.ru

DRESVYANIN Pavel, student of Institute of Radioelectronics and Information Technologies of Ural Federal University. 620002, Russia, Yekaterinburg, 32 Mira Street. E-mail: pdresvyanin@bk.ru

SAFIULLIN Nikolai, Candidate of Technical Sciences, Docent of Department of Information Technologies and Automation, Institute of Radioelectronics and Information Technologies of Ural Federal University. 620002, Russia, Yekaterinburg, 32 Mira Street. E-mail: n.t.safiullin@urfu.ru

PORSHNEV Sergey, Doctor of Technical Sciences, Professor, Director of Scientific-Educational Center for «Information Security», Institute of Radioelectronics and Information Technologies of Ural Federal University. 620002, Russia, Yekaterinburg, 32 Mira Street. E-mail: s.v.porshnev@urfu.ru