НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2018 Том 18 № 4 ISSN 2226-1494 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July–August 2018 Vol. 18 No 4 ISSN 2226-1494 http://ntv.ifmo.ru/en

УДК 004.93, 57.087.1

ДЕТЕКТИРОВАНИЕ АТАК НА ГОЛОСОВЫЕ БИОМЕТРИЧЕСКИЕ СИСТЕМЫ В ТЕЛЕФОННОМ КАНАЛЕ

Г.М. Лаврентьева[a,b]

[a] ООО «ЦРТ-инновации», Санкт-Петербург, 196084, Российская Федерация
[b] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
Адрес для переписки: lavrentyeva@speechpro.com

Информация о статье
Поступила в редакцию 01.06.18, принята к печати 25.06.18
doi: 10.17586/2226-1494-2018-18-4-663-668
Язык статьи – русский

Ссылка для цитирования: Лаврентьева Г.М. Детектирование атак на голосовые биометрические системы в телефонном канале // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 4. С. 663–668. doi: 10.17586/2226-1494-2018-18-4-663-668

Аннотация

Предмет исследования. Исследована проблема детектирования атак на голосовые биометрические системы (спуфинг-атак) в телефонном канале. На сегодняшний день детектирование атак на голосовые биометрические системы является приоритетным направлением в области аутентификации диктора по голосу. Результаты конкурса по детектированию спуфинг-атак Automatic Speaker Verification Spoofing and Countermeasures Challenge 2015 и 2017 годов подтвердили высокие перспективы в детектировании неизвестных заранее типов атак в микрофонном канале. Однако аналогичная задача в телефонном канале остается крайне актуальной, например, в банковской сфере. **Метод.** Исследован подход на основе глубоких нейронных сетей для решения описанной задачи, в частности конволюционных нейронных сетей с Max-Feature-Map активационной функцией. **Основные результаты.** Эксперименты, проведенные в рамках этого исследования на реальных телефонных атаках, показали недостаточную эффективность систем, обученных на данных с эмулированным телефонным каналом, вследствие чего была собрана база реальных атак в телефонном канале. Лучшая система продемонстрировала ошибку EER, равную 1,5%, на подмножестве атак повторного воспроизведения, 1,7% на атаках преобразования голоса и 2,8% на атаках, использующих синтезированный голос. Тем не менее, эксперименты показывают необходимость расширения обучающей выборки на различные условия записи, в силу влияния большого количества факторов на канал связи. **Практическая значимость.** Результаты работы могут найти применение в области голосовой биометрии. Представленные методы могут быть использованы в системах автоматической верификации и идентификации дикторов по голосу для детектирования атак с целью взлома.

Ключевые слова
детектирование анти-спуфинг, изменение канала связи, CNN

Благодарности
Работа выполнена в рамках темы ПНИЭР «Разработка технологии автоматической бимодальной верификации по лицу и голосу с защитой от использования подложных биометрических образцов» при финансовой поддержке Министерства образования и науки Российской Федерации по соглашению о предоставлении субсидии №14.578.21.0189 от 03.10.2016 RFMEFI57816X0189.

DETECTION OF SPOOFING ATTACKS ON SPEAKER VERIFICATION SYSTEMS IN TELEPHONE CHANNEL

G.M. Lavrentyeva[a,b]

[a] STC-innovations Ltd., Saint Petersburg, 196084, Russian Federation
[b] ITMO University, Saint Petersburg, 197101, Russian Federation
Corresponding author: lavrentyeva@speecpro.com

Article info
Received 01.06.18, accepted 25.06.18
doi: 10.17586/2226-1494-2018-18-4-663-668
Article in Russian

For citation: Lavrentyeva G.M. Detection of spoofing attacks on speaker verification systems in telephone channel. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 4, pp. 663–668 (in Russian). doi: 10.17586/2226-1494-2018-18-4-663-668

Abstract

Subject of Research. The present paper is devoted to the attacks detection problem on voice biometric systems (spoofing-attacks) in telephone channel. Nowadays, spoofing detection is under the high interest in the field of voice speaker authentication. The results of the Automatic Speaker Verification Spoofing and Countermeasures Challenge in 2015 and 2017 dedicated to isolated task of spoofing detection confirmed the high perspectives in detection of unknown types of attacks in microphone channel. However, similar task in telephone channel remains extremely relevant, for example, in the banking sector. **Method.** The aim of the work was to study the applicability of deep learning approach for described problem solution, in particular, convolutional neural networks with the Max-Feature-Map activation function. **Main Results.** The experiments performed for real telephone attacks showed insufficient efficiency of the systems trained on data with emulated telephone channel. That is why, the database of real spoofing attacks in telephone channel was collected. The best system demonstrated 1.5% equal error rate (EER) on a subset of replay attacks, 1.7% for voice conversion attacks, and 2.8% for attacks with voice synthesis. Experiments show the need to consider different recording conditions, due to the great number of factors that have the influence on the channel. **Practical Relevance.** The results of the work can be applied in the field of voice biometrics. The presented methods can be used in systems of automatic speaker verification and identification for detection of spoofing attacks on them.

Keywords

spoofing detection, channel variation, CNN

Acknowledgements

The study was performed in the framework of the research project for applied research and experimental designs "Development of technology for automatic bimodal face and voice verification with protection against the use of false biometric samples". This work was financially supported by the Ministry of Education and Science of the Russian Federation, Contract 14.578.21.0189 dated 3/10/2016 (ID RFMEFI57816X0189).

Введение

В настоящее время системы автоматической идентификации и верификации диктора предлагают удобный, надежный и экономичный метод аутентификации пользователей. Однако атаки злоумышленников с целью получения доступа к защищаемой информации, иначе называемые спуфингом, могут подорвать доверие и создать сложности в эксплуатации этих систем. Особую угрозу представляют собой атаки на устройства ввода в биометрической системе распознавания диктора (направленные спуфинг-атаки), т.е. когда злоумышленник пытается представить себя как зарегистрированный в системе пользователь. Такие атаки проще в реализации и не требуют от злоумышленников специального оборудования и знаний о системе. К известным типам направленных спуфинг относятся: имперсонализация [1], синтез, преобразование речи и повторное воспроизведение [2]. Последние три представляют собой наибольшую опасность [3]. Результаты конкурсов ASVspoof 2015 и 2017 года [4], нацеленные на усовершенствование методов детектирования атак в микрофонном канале, продемонстрировали большой потенциал современных методов в обнаружении таких атак. Однако аналогичная задача в телефонном канале является более сложной в силу того, что меньшая ширина пропускаемой полосы частот, кодирование сигнала, потеря пакетов и другие нелинейные искажения могут значительно повлиять на способность детектирования спуфинга. Основной целью настоящей работы является исследование эффективности перспективного подхода конволюционных нейронных сетей (CNN) для решения задачи детектирования спуфинг-атак в телефонном канале. Успех подхода на основе CNN, продемонстрированного в работе [5] для решения задачи детектирования спуфинг-атак повторного воспроизведения в микрофонном канале, а также эффективное использование CNN для классификации видео [6], изображений [7, 8], распознавания лиц [9], а также детектирования атак на системы лицевой биометрии [10], был мотивацией для применения такого подхода к описанной задаче.

База спуфинг-атак в телефонном канале

Исследование влияния изменения полосы пропускания частот и канальных вариаций на качество детектирования спуфинг-атак рассматривалось в работе [11]. Для этого авторы использовали базу с эмулированным телефонным каналом, полученную с помощью понижения частоты дискретизации, а также пропусканием микрофонных записей через полосные фильтры и программные GSM-кодеки. Наши системы, обученные на эмулированных данных, показали значительное снижение качества детектирования спуфинг-атак в реальном телефонном канале (до 30% EER для базы ASVspoof 2015), что подтвердило необходимость обучающей выборки, содержащей реальные спуфинг-атаки в телефонном канале, в связи с чем была собрана соответствующая база в телефонном канале.

В качестве исходных образцов записей спуфинг-атак в микрофонном канале, которые впоследствии перезаписывались в телефонном канале, использовалась база конкурса ASVspoof 2015 (только английский язык), образцы синтезированной речи, полученные средствами, доступными онлайн (библиоте-

ки и облачные сервисы): Yandex[1], Google[2], Lyrebird[3], IBM[4] на русском и английском языках и системой синтеза речи компании ЦРТ (Центр Речевых Технологий) [12] (только русский язык).

Для перезаписи базы атак в телефонном канале использовались 2 сценария (рис. 1): первый использовал систему регистрации телефонных вызовов и речевых сообщений Незабудка II [13], второй – систему голосового оповещения по телефонным линиям Рупор [14]. В обоих случаях фонограммы со спуфингом проигрывались на компьютере и по кабелю 3Jack-4Jack передавались на мобильное устройство, используемое в качестве микрофона. В это же время мобильное устройство совершало звонок либо на стационарный телефон, который был подключен к системе Незабудка II, либо на один из каналов системы Рупор. В обоих случаях конечные системы записывали входной сигнал, прошедший через все необходимые каналы связи и кодеки, тем самым получая образец голосовой атаки в телефонном канале. Для обеспечения вариативности каналов использовались 2 различных телефона (Samsung Galaxy S4, Xiaomi Redmi 4A) и 2 оператора связи (Megafon, МТС). Общее количество собранных образцов спуфинг-атак представлено в табл. 1. В качестве образцов записей реальной речи использовалась речевая база данных конкурса NIST [15].

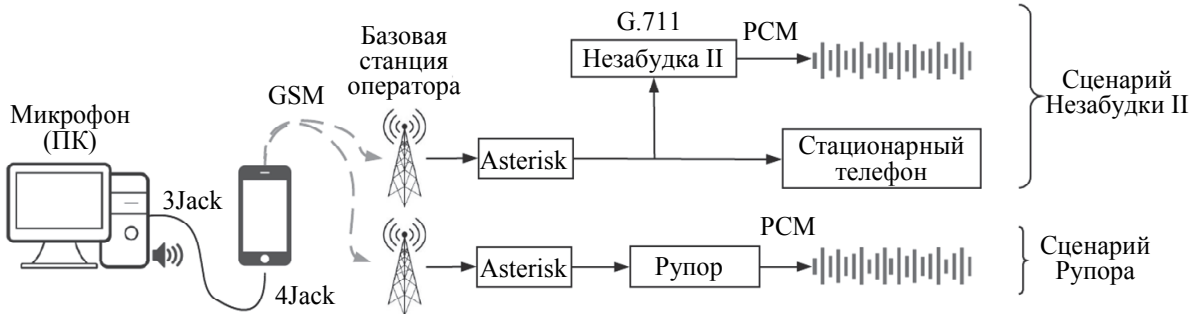

Рис. 1. Схема сбора баз

Тип атак	**Язык**	**Микрофонный канал**	**Записанные с помощью системы Незабудка II**	**Записанные с помощью системы Рупор**
ASVpoof2015 spoof	Английский	246500 / 177,1	41162 / 361,6	95229 / 856,6
ASVspoof2015 genuine	Английский	16651 / 160,2	5573 / 51,2	–
iSpeech	Английский	155 / 5,47	–	–
IBM	Английский	14228 / 203,4	–	1405 / 19,9
	Русский	9260 / 217,0	–	636 / 15,1
Zamzar	Английский	11 / 210,8	–	–
STC	Русский	1035 / 523,1	–	1120 / 48,4
Yandex	Английский	5112 / 236,3	77 / 6,2	931 / 53,4
	Русский	3716 / 201,7	39 / 3,3	1001 / 56,8
Google	Английский	5572 / 314,8	73 / 6,6	566 / 46,0
	Русский	3727 / 240,6	32 / 3,3	606 / 48,7
Lyrebird	Английский	3600 / 95,9	–	60 / 1,64
RSR phrases	Английский	15000 / 150,7	–	3369 / 29,9
RedDots2015	Английский	15305 / 142,8	–	3193 / 30

Таблица 1. Количество образцов спуфинг-атак в микрофонном и телефонном канале (количество файлов / суммарная длительность в часах)

[1] Технология синтеза Yandex [Электронный ресурс] // URL: https://tech.yandex.ru/speechkit/mobilesdk/doc/intro/overview/concepts/tts-overview-technology-docpage/
[2] Технология синтеза Google [Электронный ресурс] // URL: https://github.com/pndurette/gTTS
[3] Технология синтеза Lyrebird [Электронный ресурс] // URL: https://lyrebird.ai/
[4] Технология синтеза IBM [Электронный ресурс] // URL: https://www.research.ibm.com/tts/

Описание системы

Задача детектирования голосовых спуфинг-атак может быть сведена к детектированию локальных спектральных артефактов, присутствующих в поддельной или воспроизведенной речи и отличающих ее от реальной речи. Для детектирования атаки повторного воспроизведения в микрофонном канале в работе [5] уже использовались в качестве экстрактора высокоуровневых признаков сверточные нейронные сети с Max-Feature-Map (MFM) активационными функциями – Light CNN (LCNN). MFM представляет собой расширение активационной функции максимального выхода, которая берет максимум из двух входов. Использование такой активационной функции позволило в значительной степени уменьшить архитектуру CNN [16]. В отличие от обычных пороговых функций активации, MFM подавляет нейрон конкурентным отношением, и, таким образом, MFM играет роль селектора признаков. В качестве входных данных использовалось частотно-временное представление речевых сигналов в виде спектрограмм. При этом использовался нормализованный спектр, полученный на основе быстрого преобразования Фурье.

Так как данные, подаваемые на вход нейронной сети, должны быть в единой частотно-временной форме, для унификации применялась оконная обработка с окном фиксированного размера, т.е. каждое окно спектрограммы обрабатывалось нейронной сетью независимо.

В настоящей работе также был использован подход на основе LCNN с некоторыми модификациями. В отличие от микрофонного канала, в телефонном канале паузы содержат неинформативные шумы, тональные помехи и др. Предварительная отбраковка таких участков позволяет повысить эффективность детектирования спуфинг-атак. Для этого в качестве предварительной обработки применялся детектор речевой активности на основе выделителя основного тона [17], после чего оконная функция применялась только к речевым сегментам.

Проведенные эксперименты показали, что для решения задачи детектирования спуфинга в телефонном канале можно в 4 раза ускорить оригинальную систему из [5] за счет уменьшения размеров сверточных слоев в 4 раза, не теряя при этом в качестве детектирования спуфинг-атак. На рис. 2 представлена архитектура используемой сети.

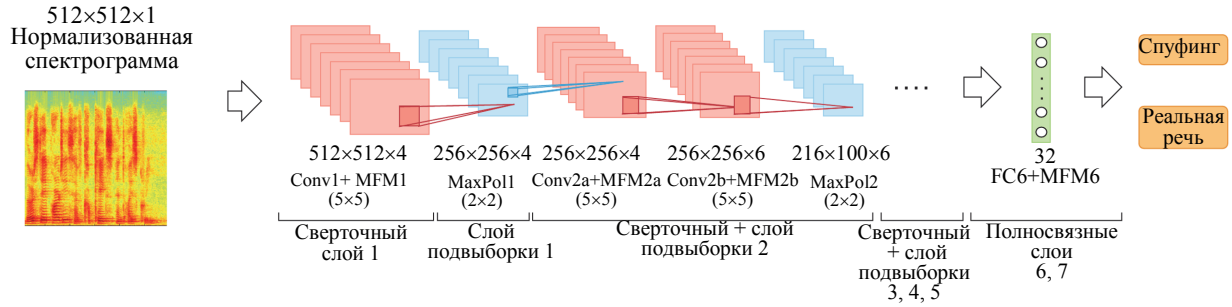

Рис. 2. Архитектура LCNN для выделения высокоуровневых признаков

Аналогично системе из [5] высокоуровневые признаки извлекались из предпоследнего полносвязного слоя, после чего подавались на вход простому классификатору. В данной работе использовался стандартный двухклассовый классификатор на основе смеси гауссовых распределений (Gaussian Mixture Models, GMM) аналогично тому, как это было использовано в [4]. Для класса реальной речи и класса спуфинга были обучены индивидуальные GMM-модели с помощью EM-алгоритма (Expectation Maximization) [18]. На этапе классификации для каждого высказывания из GMM-моделей были получены оценки правдоподобия, а конечный результат был рассчитан как логарифм отношения правдоподобия:

$$\Lambda(X) = \log L(X/\theta) - \log L(X/\gamma),$$

где X – последовательность векторов признаков тестового высказывания, L соотносится с функцией правдоподобия, θ и γ представляют собой GMM для естественной и поддельной речи.

Результаты экспериментов

Результаты экспериментов на собранных речевых базах представлены в табл. 2. Для оценки качества детектирования атак использовалась равновероятностная ошибка распознавания первого и второго рода EER. Результаты экспериментов показали высокую эффективность использования подхода на основе нейронных сетей для решения задачи детектирования спуфинг-атак в телефонном канале.

Эксперименты на базах синтезированной речи, полученных с помощью средств, доступных онлайн – Google, Yandex и IBM для английского и русского языка, демонстрируют различное качество детектирования спуфинга. Так, эффективность детектирования записей на английском языке, полученных с помощью систем синтеза Google и IBM, практически вдвое выше эффективности детектирования атак на русском языке. Противоположная ситуация наблюдается на образцах атак, полученных с помощью системы синтеза от Yandex. Принимая во внимание, что для всех систем синтеза объемы русского и английского обучающего материала были одинаковы, такие результаты могут быть объяснены различием в ка-

честве синтеза речи различных компаний. Более высокое качество синтеза русской речи от Yandex может быть обусловлено целевым рынком, объемами баз и тем, что русский язык является родным для основной части разработчиков.

Типы атак	База воспроизводимых записей речи	Язык	Система записи	EER, %
Повторное воспроизведение	Genuine часть базы ASVspoof 2015	Английский	Незабудка II	1,5
Синтез	Google	Английский	Незабудка II Рупор	1,33
		Русский		0,74
	Yandex	Английский		0,21
		Русский		1,27
	IBM	Английский		4,49
	IBM	Русский		1,95
	ASVspoof 2015	Английский	Незабудка II	3,2
			Рупор	0,98
Преобразование голоса	ASVspoof 2015	Английский	Незабудка II	3,61
			Рупор	1,38

Таблица 2. Результаты экспериментов

Интерес также представляют оценки качества детектирования атак типа синтеза и преобразования речи из базы ASVspoof 2015 для различных сценариев записи атак. Оба сценария использовались для записи обучающей базы в равных пропорциях. Разные показатели EER для систем Незабудка II и Рупор в обоих случаях свидетельствуют о сильном влиянии особенностей канала записи на способность детектирования спуфинга в них.

Заключение

В настоящей работе рассматривалась применимость подхода на основе глубоких нейронных сетей для решения задачи детектирования спуфинг-атак на системы верификации диктора по голосу в телефонном канале. Для этого была собрана обучающая и тестовая база спуфинг-атак на основе доступных систем синтеза речи и речевой базы конкурса ASVspoof 2015. Проведенные на собранной базе эксперименты показали высокую эффективность предложенного подхода. Лучшая по результатам экспериментов система показала ошибку EER меньше 5% для всех типов спуфинг-атак. Однако стоит отметить, что полученная в работе система была обучена и протестирована на ограниченном наборе характеристик канала. Полученные результаты подтверждают необходимость дальнейшего исследования влияния канала на способность детектирования спуфинг-атак.

Описанные системы детектирования спуфинг-атак могут быть использованы совместно с системами голосовой верификации пользователей. Предложенная система детектирования спуфинг-атак крайне актуальна в различных приложениях контроля доступа, например, в банковской сфере, силовых ведомствах, государственных структурах [19].

Литература

1. Hautamki R., Kinnunen T., Hautamki V., Laukkanen A.-M. Automatic versus human speaker verification: the case of voice mimicry // Speech Communication. 2015. V. 72. P. 13–31. doi: 10.1016/j.specom.2015.05.002
2. Evans N., Kinnunen T., Yamagishi J. Spoofing and countermeasures for automatic speaker verification // Proc. of Interspeech. Lyon, France, 2013. P. 925–929.
3. Wu Z., Evans N., Kinnunen T., Yamagishi J., Alegre F., Li H. Spoofing and countermeasures for speaker verification: a survey // Speech Communication. 2015. V. 66. P. 130–153. doi: 10.1016/j.specom.2014.10.005
4. Wu Z., Yamagishi J., Kinnunen T., Hanilci C., Sahidullah M., Sizov A., Evans N., Todisco M., Delgado H. ASVspoof: the automatic speaker verification spoofing and countermeasures challenge // IEEE Journal on Selected Topics in Signal Processing. 2017. V. 11. N 4. P. 588–604. doi: 10.1109/JSTSP.2017.2671435

References

1. Hautamki R., Kinnunen T., Hautamki V., Laukkanen A.-M. Automatic versus human speaker verification: the case of voice mimicry. *Speech Communication*, 2015, vol. 72, pp. 13–31. doi: 10.1016/j.specom.2015.05.002
2. Evans N., Kinnunen T., Yamagishi J. Spoofing and countermeasures for automatic speaker verification. *Proc. of Interspeech*. Lyon, France, 2013, pp. 925–929.
3. Wu Z., Evans N., Kinnunen T., Yamagishi J., Alegre F., Li H. Spoofing and countermeasures for speaker verification: a survey. *Speech Communication*, 2015, vol. 66, pp. 130–153. doi: 10.1016/j.specom.2014.10.005
4. Wu Z., Yamagishi J., Kinnunen T., Hanilci C., Sahidullah M., Sizov A., Evans N., Todisco M., Delgado H. ASVspoof: the automatic speaker verification spoofing and countermeasures challenge. *IEEE Journal on Selected Topics in Signal Processing*, 2017, vol. 11, no. 4, pp. 588–604. doi: 10.1109/JSTSP.2017.2671435

5. Lavrentyeva G., Novoselov S., Malykh E., Kozlov A., Kudashev O., Shchemelinin V. Audio replay attack detection with deep learning frameworks // Proc. of Interspeech. Stockholm, Sweden, 2017. P. 82–86. doi: 10.21437/Interspeech.2017-360
6. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., Fei-Fei L. Large-scale video classification with convolutional neural networks // Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. Columbus, USA, 2014. P. 1725–1732. doi: 10.1109/CVPR.2014.223
7. Bengio Y., Courville A., Vincent P. Representation learning: a review and new perspectives // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013. V. 35. N 8. P. 1798–1828. doi: 10.1109/TPAMI.2013.50
8. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks // Advances Inneural Information Processing Systems. Lake Tahoe, USA, 2012. P. 1097–1105.
9. Taigman Y., Yang M., Ranzato M., Wolf L. DeepFace: closing the gap to human-level performance in face verification // Proc. of IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014. P. 1701–1708. doi: 10.1109/CVPR.2014.220
10. Волкова С.С., Матвеев Ю.Н. Применение сверточных нейронных сетей для решения задачи противодействия атаке спуфинга в системах лицевой биометрии // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 4. С. 702–710. doi: 10.17586/2226-1494-2017-17-4-702-710
11. Delgado H., Todisco M., Evans N., Sahidullah M., Liu W.M., Alegre F., Kinnunen T., Fauve B. Impact of bandwidth and channel variation on presentation attack detection for speaker verification // Lecture Notes in Informatics. Darmstadt, Germany, 2017. Art. 8053510. doi: 10.23919/BIOSIG.2017.8053510
12. Chistikov P., Zakharov D., Talanov A. Improving speech synthesis quality for voices created from an audio book database // Lecture Notes in Computer Science. 2014. V. 8773. P. 276–283.
13. Многоканальная система регистрации телефонных вызовов и речевых сообщений Незабудка II [Электронный ресурс]. URL: https://www.speechpro.ru/product/sistemy-zapisi-telefonnykh-razgovorov/nezabudka-2, своб. Яз. рус. (дата обращения 05.06.2018)
14. Многоканальная система автоматического оповещения абонентов по телефонным линиям Рупор [Электронный ресурс]. URL: https://www.speechpro.ru/product/sistemy-rechevogo-opovesheniya/rupor, своб. Яз. рус. (дата обращения 05.06.2018)
15. NIST Speaker Recognition Evaluation 2012 Database [Электронный ресурс]. URL: https://www.nist.gov/itl/iad/mig/sre12-results, своб. Яз. рус. (дата обращения 05.06.2018)
16. Wu X., He R., Sun Z., Tan T. A light CNN for deep face representation with noisy labels // IEEE Journal of Selected Topics in Signal Processing. 2018. V. 13. N 11. P. 2884–2896. doi: 10.1109/TIFS.2018.2833032
17. Симончик К.К., Галинина О.С., Капустин А.И. Алгоритм обнаружения речевой активности на основе статистик основного тона в задаче распознавания диктора // Научно-технические ведомости СПбГПУ. 2010. № 4(103). С. 18–23.
18. Markov K., Nakagawa S. Discriminative training of GMM using a modified EM algorithm for speaker recognition // Proc. of International Speech Communication Association. Sydney, Australia, 1998.
19. Дырмовский Д.В., Коваль С.Л., Хитров М.В. Концепция системы национального фоноучета и голосового биометрического поиска // Известия вузов. Приборостроение. 2014. Т. 57. № 2. С. 63–70.

Авторы

Лаврентьева Галина Михайловна – научный сотрудник, ООО «ЦРТ-инновации», Санкт-Петербург, 196084, Российская Федерация; аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 56938815200, ORCID ID: 0000-0001-9474-098X, lavrentyeva@speechpro.com

5. Lavrentyeva G., Novoselov S., Malykh E., Kozlov A., Kudashev O., Shchemelinin V. Audio replay attack detection with deep learning frameworks. *Proc. of Interspeech*. Stockholm, Sweden, 2017, pp. 82–86. doi: 10.21437/Interspeech.2017-360
6. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., Fei-Fei L. Large-scale video classification with convolutional neural networks. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*. Columbus, USA, 2014, pp. 1725–1732. doi: 10.1109/CVPR.2014.223
7. Bengio Y., Courville A., Vincent P. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, no. 8, pp. 1798–1828. doi: 10.1109/TPAMI.2013.50
8. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. *Advances Inneural Information Processing Systems*. Lake Tahoe, USA, 2012, pp. 1097–1105.
9. Taigman Y., Yang M., Ranzato M., Wolf L. DeepFace: closing the gap to human-level performance in face verification. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014, pp. 1701–1708. doi: 10.1109/CVPR.2014.220
10. Volkova S.S., Matveev Yu.N. Convolutional neural networks for face anti-spoofing. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 4, pp. 702–710 (in Russian). doi: 10.17586/2226-1494-2017-17-4-702-710
11. Delgado H., Todisco M., Evans N., Sahidullah M., Liu W.M., Alegre F., Kinnunen T., Fauve B. Impact of bandwidth and channel variation on presentation attack detection for speaker verification. *Lecture Notes in Informatics*. Darmstadt, Germany, 2017, art. 8053510. doi: 10.23919/BIOSIG.2017.8053510
12. Chistikov P., Zakharov D., Talanov A. Improving speech synthesis quality for voices created from an audio book database. *Lecture Notes in Computer Science*, 2014, vol. 8773, pp. 276–283.
13. *Multi-channel system for registering telephone calls and voice messages Nezabudka II*. Available at: https://www.speechpro.ru/product/sistemy-zapisi-telefonnykh-razgovorov/nezabudka-2 (accessed 05.06.2018).
14. *Multi-channel system of automatic notification of subscribers over telephone lines Rupor*. Available at: https://www.speechpro.ru/product/sistemy-rechevogo-opovesheniya/rupor (accessed 05.06.2018).
15. *NIST Speaker Recognition Evaluation 2012 Database*. Available at: https://www.nist.gov/itl/iad/mig/sre12-results (accessed 05.06.2018).
16. Wu X., He R., Sun Z., Tan T. A light CNN for deep face representation with noisy labels. *IEEE Journal of Selected Topics in Signal Processing*, 2018, vol. 13, no. 11, pp. 2884–2896. doi: 10.1109/TIFS.2018.2833032
17. Simonchik K.K., Galinina O.S., Kapustin A.I Algorithm for detecting speech activity based on the pitch statistics in the task of recognizing the speaker. *Nauchno-Tekhnicheskie Vedomosti SPbGPU*, 2010, no. 4, pp. 18–23. (in Russian)
18. Markov K., Nakagawa S. Discriminative training of GMM using a modified EM algorithm for speaker recognition. *Proc. of International Speech Communication Association*. Sydney, Australia, 1998.
19. Dyrmovsky D.V., Koval S.L., Khitrov M.V. Concept of the national voice accounting and voice biometric search system. *Journal of Instrument Engineering*, 2014, vol. 57, no. 2, pp. 63–70. (in Russian)

Authors

Galina M. Lavrentyeva – Scientific researcher, STC-innovations Ltd., Saint Petersburg, 196084, Russian Federation; postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 56938815200, ORCID ID: 0000-0001-9474-098X, lavrentyeva@speechpro.com