



## ПРОГРАМНО-ТЕХНИЧНИ КОМПЛЕКСИ

В.И. СОЛОВЬЕВ, О.В. РЫБАЛЬСКИЙ, В.В. ЖУРАВЕЛЬ

УДК 621.317+681.849

### ОБОСНОВАНИЕ ПРИНЦИПИАЛЬНОЙ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПОСТРОЕНИЯ СИСТЕМЫ ВЫЯВЛЕНИЯ СЛЕДОВ ЦИФРОВОЙ ОБРАБОТКИ ФОНОГРАММ

**Аннотация.** На модели нейронной сети глубокого обучения обоснована и проверена принципиальная пригодность данной сети для создания высокоэффективного экспертного инструментария, предназначенного для выявления следов цифровой обработки в фонограммах. Эксперимент проведен на большом объеме (более 100000) экспериментальных фрагментов необработанных пауз фонограмм и пауз со следами обработки, полученных в автоматическом режиме. Найденные зависимости показали, что при пороге вероятности свыше 0.55 правильной бинарной классификации пауз возникает возможность построения высокоэффективного инструментария экспертизы.

**Ключевые слова:** нейронная сеть глубокого обучения, точки монтажа, цифровая обработка фонограммы, цифровая фонограмма, цифровой монтаж, экспертиза.

Одной из наиболее важных проблем экспертизы материалов и аппаратуры звукозаписи является проверка отсутствия (или наличия) следов цифровой обработки в фонограммах. Она возникла с началом широкого использования аппаратуры цифровой звукозаписи (АЦЗЗ). Существует несколько направлений создания инструментария для решения этой проблемы.

Одно из них — развиваемое во Франции, Великобритании, России, Бразилии, Польше, Испании и Румынии, основано на выделении и исследовании регулярных спектральных компонентов из записанных сигналов [1–14]. При этом в большинстве инструментов, разработанных на основе этого направления, выделяют спектральные составляющие, образующиеся за счет электромагнитных сетевых наводок на аппаратуру записи. Однако в современной АЦЗЗ они, как правило, отсутствуют в записываемых цифровых сигналах. Кроме того, весь такой инструментарий для выделения регулярных спектральных компонентов использует кратковременное преобразование Фурье, разрешающая способность которого не позволяет выявлять следы цифровой обработки фонограмм ввиду их чрезвычайной малости [15]. Поэтому эффективность такого инструментария не обеспечивает достаточной достоверности результатов экспертизы.

При этом для большинства предложенных решений обязательно проведение сравнительных исследований параметров образцовых (экспериментальных) и спорных (представленных на экспертизу) фонограмм. Поэтому для записи образцовых фонограмм требуется проводить экспертизу аппаратуры, на которой записывались спорные фонограммы, что часто вообще не выполнимо ввиду ряда специфических обстоятельств [16].

© В.И. Соловьев, О.В. Рыбальский, В.В. Журавель, 2020

Кроме того, широко используемый в странах ЕС метод, основанный на проверке частоты (или фазы) сигнала сетевой наводки, зафиксированного в фонограмме, имеет ряд существенных недостатков. Часто он не позволяет выявлять следы цифровой обработки (монтажа), выполненной методом вырезания и перестановки (с использованием пауз в речевых сигналах) фрагментов одной фонограммы [17]. Такие следы наиболее сложны для обнаружения экспертизой.

Инструментарий выявления следов внешних вмешательств пытались построить на принципе рассмотрения цифровой фонограммы как цифрового файла со своими атрибутами, по метаданным которого проводится экспертный анализ [18]. Это направление ввиду своей ограниченности не обеспечивает достоверности экспертных решений, принимаемых на основании его использования, поэтому он широко не применяется в экспертизе и иногда используется как вспомогательный.

Разработка еще одного направления создания такого инструментария началась с теоретического осмысления процессов, происходящих при записи и цифровой обработке фонограмм [19]. Устанавливались источники возникновения паразитных явлений, происходящих при цифровой записи, общие для всех видов АЦЗЗ, что позволяло установить строго индивидуальный характер искажений, вносимых в сигналы каждым конкретным экземпляром такой аппаратуры. Это направление экспертизы основано на строгой индивидуальности паразитных явлений в АЦЗЗ, фиксируемых в собственных шумах фонограмм [15, 19]. Оно развивается на протяжении всех последних лет, что обусловлено перспективой создания эффективного экспертного инструментария для выявления следов цифровой обработки в фонограммах. В процессе его исследования были разработаны основы теории выявления следов цифровой обработки фонограмм, предложены пути и способы решения проблемы, обеспечивающие высокую степень эффективности обнаружения монтажа в фонограммах [20].

В частности, в ряде работ авторов рассматривались некоторые модели проявления в сигналах следов цифровой обработки фонограмм, выполненной способом вырезания и перестановки фрагментов [21]. Затем экспериментально доказывалось, что следы такой обработки проявляются в виде изменений фрактального состава монтируемых сигналов. Отмечалось, что хотя такие проявления и имеются, но выделить их в виде конкретных признаков невозможно без проведения обширных исследований, для чего необходимо создать специальное программное обеспечение [22, 23]. Дальнейший поиск направления создания таких программ показал, что наиболее перспективно использование нейронных сетей глубокого обучения [24].

Отметим, что до настоящего времени в публикациях по этой тематике практически не приводятся величины оценки эффективности различных методов выявления следов цифровой обработки в фонограммах. Это не позволяет качественно проанализировать характеристики таких систем.

Целью настоящей работы является экспериментальная проверка возможности создания эффективной системы, построенной на основе нейронной сети глубокого обучения, предназначенной для выявления следов цифровой обработки фонограмм.

В настоящее время не существует научно-обоснованного подхода, который позволил бы применять единую методологию выявления следов цифровой обработки в фонограммах. Отметим, что подобная методология должна базироваться на большом объеме экспериментальных данных. Однако апробация экспертами любого предлагаемого подхода с учетом большого массива статистических данных в режиме «ручного исследования» практически нереализуема. Такая проверка необходима для выяснения, пригодно ли в принципе выбранное направление. Суть этой проверки состоит в исследовании способности таких систем выявлять

смонтированные паузы. Для этого в звуковом редакторе из фонограмм, записанных на различной АЦЗЗ, в «ручном режиме» проводился предварительный отбор пауз длительностью от 100 мс до нескольких секунд. Отбирались приблизительно 1000 таких пауз (при частоте дискретизации 44.1 кГц), которые и составили исходную базу эксперимента. Из отобранных пауз в автоматическом (компьютерном) режиме формировался массив фрагментов пауз. Для этого они сканировались окном длительностью 15–30 мс с интервалом 1 мс и вырезались фрагменты данной длительности. Таким образом, формировались массивы фрагментов пауз различной длительности размерностью приблизительно 100000. Для формирования массива фрагментов пауз со следами цифровой обработки случайным образом выбранные паузы делились на части в произвольном соотношении. После чего из полученных частей формировались новые паузы с зафиксированным местом монтажа. Затем сканированием этих пауз временным окном из зоны локализации монтажа вырезались фрагменты пауз со следами обработки одинаковой длительности. Размерность массива фрагментов пауз с монтажом составляла приблизительно 100000.

Упомянутые массивы фрагментов пауз с обработкой и без нее явились исходными массивами (dataset) для обучения нейронной сети.

Исследования в области речевых технологий и обработки аудиоинформации показали высокую эффективность полносвязных сетей глубокого обучения [25]. Поэтому для эксперимента разрабатывалась программная система, реализующая структурную модель нейронной сети. Она базировалась на библиотеке глубокого обучения нейронных сетей keras (backend tensorflow). Применялась полносвязная нейронная сеть до 50 нейронных слоев с прореживанием (dropout) и пакетной нормализацией. Решалась задача бинарной классификации на массиве dataset. В проиллюстрированных далее экспериментах фрагменты массивов пауз длительностью 20 мс имели размерность 882.

Общий массив всех фрагментов пауз делился на тренировочный и тестовый. В качестве критерия применялась наибольшая эффективность, достигаемая на тестовом массиве в процессе обучения.

В общем случае эта эффективность определяется наибольшей величиной вероятности правильной классификации фрагментов с обработкой и без нее.

Процесс проведения эксперимента с использованием нейронной сети глубокого обучения проиллюстрирован на рис. 1, 2. Результат, достигнутый сетью на обучающем массиве, показан на рис. 1, здесь эффективность, достигнутая на базе обучения за 575 эпох, составляет 0.85. Такой результат получен в процессе обучения на компьютере с графическим адаптером (GPU). На тестовом массиве максимальная эффективность составляет 0.76 (см. рис. 1).

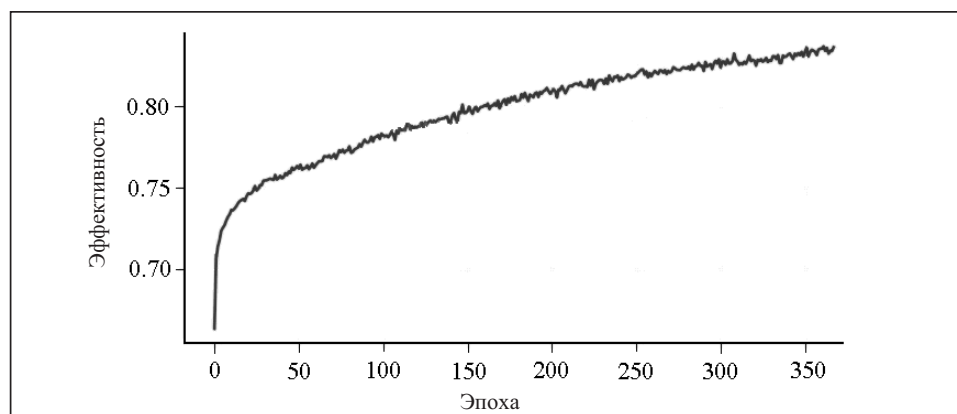


Рис. 1. График эффективности обучения

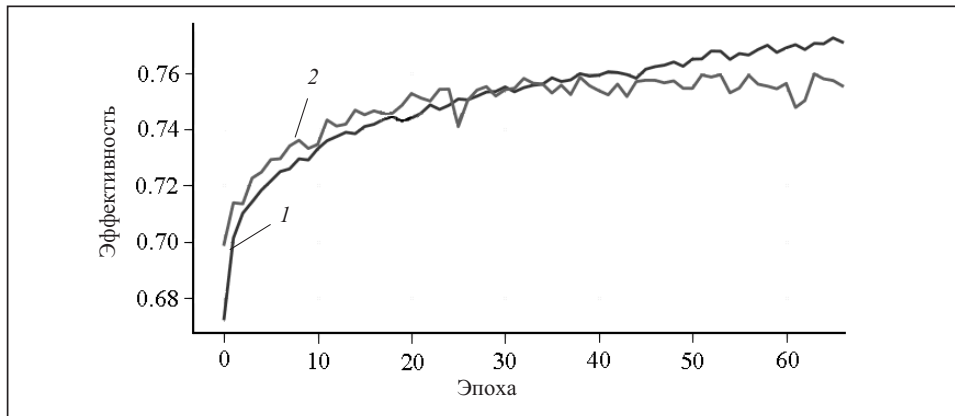


Рис. 2. Графики эффективности обучения: кривая 1 — тренировочный массив, кривая 2 — тестовый массив

На рис. 2 показаны графики для тренировочного и тестового массивов фрагментов пауз. В рамках проведенной оптимизации сети достигнута эффективность классификации 0.76. Эта вероятность, безусловно, не может являться предельной ввиду многообразия вариантов оптимизации. Так, в частности, как показывают предварительные эксперименты, существенное повышение размерности массивов фрагментов пауз и, как следствие, возможность увеличения сложности нейронной сети позволяют увеличить эффективность.

Однако в подобных задачах вряд ли можно ожидать такой же высокой эффективности классификации, как, например, в задачах классификации изображений (0.99). В то же время структура задачи позволяет предложить рациональные решения, которые можно применять в задачах выявления следов цифровой обработки в фонограммах. Ключевым моментом в решении данной задачи является введение понятия «пороговая вероятность».

В задачах классификации обученная модель позволяет получать значение величины вероятности правильной классификации. Как показывает анализ, приблизительно 30 % фрагментов пауз имеют вероятность правильной классификации ниже значения 0.55. При превышении этого порога вероятности правильной классификации резко возрастают и достигают значения 0.99 и выше. Так, на рис. 3, 4 приведены графики зависимостей вероятности правильной классификации и прогнозируемых вероятностей ошибок классификации от пороговой вероятности.

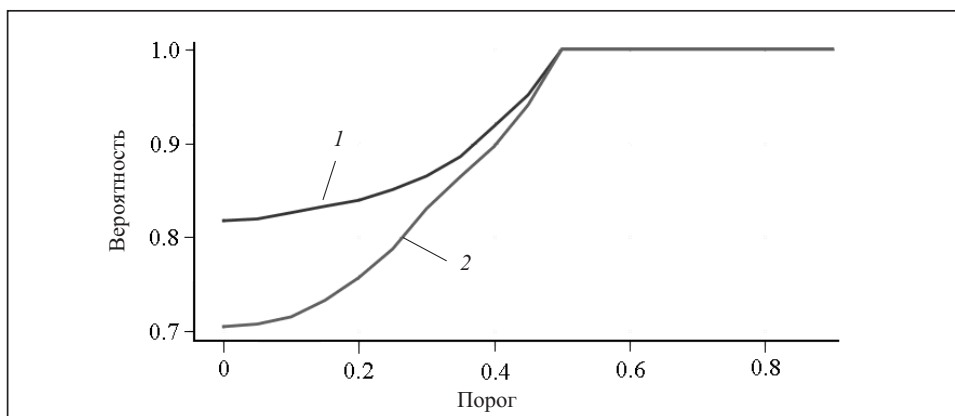


Рис. 3. Графики вероятности правильной классификации: кривая 1 — «чистые» паузы, кривая 2 — паузы с монтажом

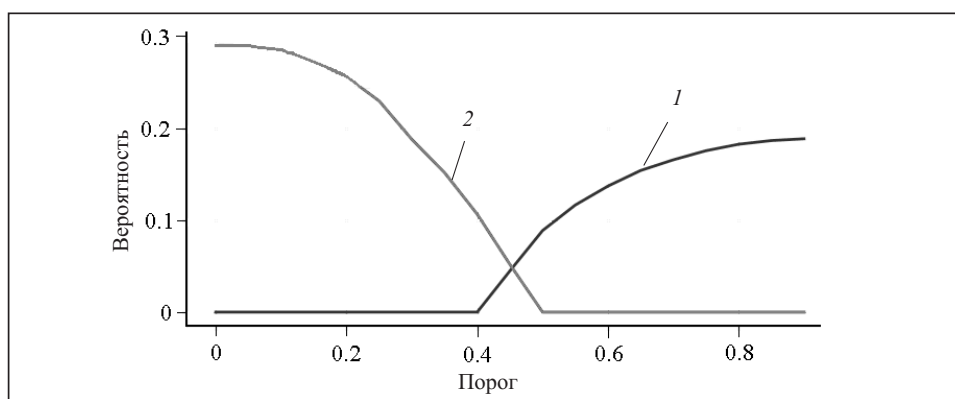


Рис. 4. Графики прогнозируемых вероятностей ошибок классификации: кривая 1 — «чистые» паузы, кривая 2 — паузы с монтажом

Отметим, что значения вероятности 0 и 1, приведенные на графиках (см. рис. 3, 4), в реальности являются приближенными к ним очень близкими значениями.

В весьма специфических задачах практического выявления следов цифрового монтажа в фонограммах можно установить высокую пороговую вероятность, например, равную 0.9. При этом вероятность правильной классификации будет не ниже 0.99. Однако на практике для этой задачи и такая вероятность недостаточна. Так, на фрагментах пауз без монтажа общей длительностью 1 с при такой вероятности правильной классификации возможны 10 ложных результатов.

Отметим, что введение пороговой вероятности при решении практических задач экспертизы соответствует пропуску фрагментов пауз с признаками следов обработки и ложных срабатываний с низкой вероятностью. В то же время это соответствует природе исследуемого объекта (пауз). Это обусловлено тем, что и в паузах без обработки, как правило, могут возникать выбросы малого уровня.

Проведенные экспериментальные исследования показали пригодность нейронных сетей глубокого обучения для построения системы выявления следов цифровой обработки в фонограммах. Перспективность применения таких сетей состоит в возможности дальнейшего усовершенствования их конфигурации, что позволяет рассчитывать на достижение высокой эффективности описанной системы. Однако методы таких усовершенствований требуют отдельного рассмотрения.

#### СПИСОК ЛИТЕРАТУРЫ

1. ENFSI working group for forensic speech and audio analysis. URL: <http://www.enfsi.eu/about-enfsi/structure/working-groups/speech-and-audio>.
2. Korycki R. Methods of time-frequency analysis in authentication of digital audio recordings. *INTL Journal of Electronics and Telecommunications*. 2010. Vol. 56, N 3. P. 257–261.
3. Nicolalde D.P., Apolinario J.A. Evaluating digital audio authenticity with spectral distances and ENF phase change. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009. P. 1417–1420.
4. Nicolalde D.P., Apolinario J.A., Biscainho L.W.P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis. *IEEE Transactions on Information Forensics and Security*. 2010. Vol. 5, Iss. 3. P. 534–543.
5. Rappaport D. Establishing a standard for digital audio authenticity: A critical analysis of tools, methodologies, and challenges. Thesis directed by Catalin Grigoras. URL: <http://aquarius.ime.br/~apolin/papers/IEEETIFS2010Daniel.pdf>.

6. Cooper A.J. An automated approach to the Electric Network Frequency (ENF) criterion — Theory and practice. *International Journal of Speech Language and the Law*. 2009. Vol. 16, N 2. P. 193–218.
7. Huijbregtse M., Geradts Z. Using the ENF criterion for determining the time of recording of short digital audio recordings. In: *Computational Forensics. IWCF 2009. Lecture Notes in Computer Science*. Geradts Z.J.M.H., Franke K.Y., Veenman C.J. (Eds). 2009. Vol. 5718. P. 116–124.
8. Catalin G. Statistical tools for multimedia forensics. *The 39th AES International Conference Audio Forensics: Practices and Challenges* (June 17, 2010), Hillerod, Denmark, 2010.
9. Jenkins C.W. An investigative approach to configuring forensic electric network frequency databases. Master's Thesis, University of Colorado Denver, 2011.
10. Brixen E. Audio metering measurements, Standards and practices. Second ed. Oxford, United Kingdom: Focal Press, 2011. 264 p.
11. Cooper A.J. Detection of copies of digital audio recordings produced using analogue interfacing. *International Journal of Speech, Language, and the Law*. 2008. Vol. 15, N 1. P. 67–95.
12. Grigoras C., Smith J.M., Jenkins C.W. Advances in ENF database configuration for forensic authentication of digital media. *131st Convention of the Audio Engineering Society* (October 20–23, 2011), New York, 2011.
13. Moon C.-B., Kim H., Kim B.M. Audio recorder identification using reduced noise features. In: *Ubiquitous Information Technologies and Applications. Lecture Notes in Electrical Engineering*. Berlin; Heidelberg: Springer, 2014. Vol. 280. P. 35–42.
14. Aggarwal R., Singh S., Roul A.K., Khanna N. Cellphone identification using noise estimates from recorded audio. In: *Communications and Signal Proc. (ICCSP), 2014 International Conference on. IEEE*. 2014, P. 1218–1222.
15. Рыбальський О.В. Застосування вейвлет-аналізу для виявлення слідів цифрової обробки аналогових і цифрових фонограм у судово-акустичній експертизі. Київ: Нац. акад. внутр. справ України, 2004. 167 с.
16. Журавель В.В. Особенности формирования фонограмм, записанных с телефонных каналов. *Сучасна спеціальна техніка*. 2015. № 4 (43). С. 26–31.
17. Рыбальский О.В. Модели нестандартной подделки цифровых фонограмм. *Ресстрація, зберігання і обробка даних*. 2003. Т. 5, № 4. С. 25–32.
18. Бобрицький С.М., Стороженко С.В. Дослідження ознак монтажу записів, виконаних цифровими записуючими пристроями. *Зб. наук. праць «Теорія та практика судової експертизи і криміналістики»*. Харків: Право, 2011. Вип. 11. С. 353–361.
19. Рыбальский О.В., Жариков Ю.Ф. Современные методы проверки аутентичности магнитных фонограмм в судебно-акустической экспертизе. Киев: Нац. акад. внутр. дел Украины, 2003. 302 с.
20. Рыбальский О.В. Основные положения теории выявления следов цифровой обработки фонограмм и особенности ее программной и методической реализации в экспертизе материалов и средств видеозвукозаписи. Ч. 1. *Захист інформації*. 2006. Т. 8, № 1 (28). С. 71–76.
21. Рыбальский О.В., Соловьев В.И., Журавель В.В. Следы монтажа в цифровых фонограммах, выполненного способом вырезания и перестановки фрагментов. *Ресстрація, зберігання і обробка даних*. 2016. Т. 18, № 1. С. 32–41.
22. Рыбальский О.В., Журавель В.В. Экспериментальное подтверждение результатов моделирования механизма возникновения идентификационных признаков монтажа в цифровых фонограммах. *Сучасні інформаційні та електронні технології: Зб. наук. пр. 17 Міжнародної науково-практичної конференції* (Одеса, 23–27 травня, 2016 р.). С. 125–126 .
23. Рыбальский О.В., Соловьев В.И., Журавель В.В. Экспериментальная проверка эффекта изменения фрактального состава сигналов при монтаже фонограммы способом вырезания и перестановки фрагментов. *Сучасна спеціальна техніка*. 2016. № 3 (46). С. 75–85.

24. Рыбальский О.В., Соловьев В.И., Журавель В.В. Основные требования к системе выявления точек цифрового монтажа в фонограммах и методология ее создания. *Информатика та математичні методи в моделюванні*. 2018. Т. 8, № 3. С. 232–237.
25. Yoshua Bengio. Deep learning. Lxmls 2015. Machine Learning Summer Shool. Lisbon, Portugal, 2015, 124 p. URL: <http://www.iro.umontreal.ca/~bengioy/dlbook/>.

Надійшла до редакції 11.04.2019

### **В.І. Соловйов, О.В. Рыбальський, В.В. Журавель**

#### **ОБҐРУНТУВАННЯ ПРИНЦИПОВОЇ ПРИДАТНОСТІ НЕЙРОННИХ МЕРЕЖ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ПОБУДОВИ СИСТЕМИ ВИЯВЛЕННЯ СЛІДІВ ЦИФРОВОГО ОБРОБЛЕННЯ ФОНОГРАМ**

**Анотація.** На моделі нейронної мережі глибокого навчання обґрунтовано і перевірено принципову придатність такої мережі для створення високоефективного експертного інструментарію, призначеного для виявлення слідів цифрового оброблення у фонограмах. Експеримент проведено на великому обсязі (більше 100000) експериментальних фрагментів необроблених пауз фонограм і пауз із слідами оброблення, отриманих в автоматичному режимі. Отримані залежності показали, що за порогу ймовірності понад 0.55 правильної бінарної класифікації пауз виникає можливість побудови високоефективного інструментарію експертизи.

**Ключові слова:** нейронна мережа глибокого навчання, точки монтажу, цифрове оброблення фонограми, цифрова фонограма, цифровий монтаж, експертиза.

### **V.I. Solovyov, O.V. Rybalskiy, V.V. Zhuravel**

#### **SUBSTANTIATING THE FUNDAMENTAL FITNESS OF DEEP LEARNING NEURON NETWORKS FOR CONSTRUCTION OF A SYSTEM FOR DETECTING TRACES OF DIGITAL TREATMENT OF PHONOGRAMS**

**Abstract.** On the model of a deep learning neuron network, the authors substantiate and verify principal applicability of such network to create a highly efficient expert tool intended to detect traces of digital treatment in phonograms. An experiment is conducted on a large volume (more than 100,000) of experimental fragments of untitled pauses of phonograms and pauses with traces of treatment, obtained in the automatic mode. The obtained dependences showed that for the probability threshold over 0.55 correct binary classification of pauses there is a possibility of constructing a highly efficient tool of examination.

**Keywords:** deep learning neuron network, points of editing, digital treatment of a phonogram, digital phonogram, digital editing, examination.

#### **Соловьев Виктор Иванович,**

кандидат техн. наук, доцент, заместитель заведующего кафедрой Восточноукраинского национального университета имени Владимира Даля, Северодонецк, e-mail: edemsvi@gmail.com.

#### **Рыбальский Олег Владимирович,**

доктор техн. наук, профессор, главный научный сотрудник, профессор кафедры Национальной академии внутренних дел, Киев, e-mail: gov\_1946@ukr.net.

#### **Журавель Вадим Васильевич,**

кандидат техн. наук, заведующий лабораторией Киевского научно-исследовательского экспертно-криминалистического центра МВД Украины, Киев, e-mail: fonoscopia@ukr.net.