

УДК 000.684

В. И. Соловьев

Восточноукраинский национальный университет
Молодежный квартал, 20-А, 91034 Луганск, Украина

Идентификация аппаратуры аудиозаписи по статистическим характеристикам аудиофайлов

Приведены результаты исследований по разработке методологии и математической модели выявления и анализа самоподобных структур в паузах аудиофайлов. Модель позволяет выделять статистические характеристики паразитных индивидуальных параметров аппаратуры, фиксируемых в аудиозаписи. Разработана модель идентификации аппаратуры аудиозаписи по статистическим характеристикам этих параметров в паузах аудиофайлов.

Ключевые слова: самоподобные структуры, паузы аудиофайлов, идентификация аппаратуры аудиозаписи.

Введение

Среди актуальных задач обработки аудиоинформации последнего десятилетия особое место занимает задача идентификации аппаратуры аудиозаписи по статистическим характеристикам ее паразитных параметров, фиксируемым в аудиофайлах [1, 2]. В работах [2, 3] показано, что отдельные узлы цифровой аппаратуры аудиозаписи могут оставлять индивидуальные «отпечатки», проявляющиеся в характеристиках шумов аудиоданных. В [3] показано, что идентификация цифровой аппаратуры аудиозаписи по индивидуальным характеристикам ее шумов, требует применения математических методов анализа с высоким разрешением по пространственно-временным параметрам. Первой технической реализацией идей, предложенных, например, в работе [3], стала программа «Академия» [4, 5]. Однако данная программа и методика ее применения имели ряд существенных недостатков, основным из которых была низкая повторяемость полученных результатов (например, сравнение между собой одного и того же файла давало некоторое различие в графиках, описывающих его параметры), что существенно подрывало доверие к результатам экспертизы [6]. Иначе говоря, программа не отвечала отдельным требованиям теории криминалистической идентификации. Кроме того, проведение экспертизы по этой методике было очень трудоемким и не обеспечивало полной проверки всего звукового файла [6]. Для устранения этих недостатков были проведены дополнительные исследования, результаты которых опубли-

кованы в ряде работ [1, 7–13]. Их результатом стала разработка нового программного обеспечения «Фрактал» и методики его применения, внедренных во всех экспертных учреждениях Украины, проводящих фоноскопические экспертизы [11].

Следует отметить, что анализ открытых источников свидетельствует о том, что других публикаций, относящихся к созданию методов и средств инструментальной идентификации цифровой аппаратуры звукозаписи, в Украине и в других странах не выявлено. Большинство обнаруженных исследований относятся к выявлению цифрового монтажа лингвистическими методами (это, в частности, относится к АНО «Судебный эксперт»), а вопросы инструментальных идентификационных исследований такой аппаратуры, отвечающих требованиям теории криминалистической идентификации, в других странах не решены.

В данной статье показана математическая модель, позволяющая решать задачи идентификации цифровой аппаратуры аудиозаписи по статистическим характеристикам паразитных параметров аудиофайлов, положенная в основу программы «Фрактал».

Постановка задачи и исследования

Рассмотрим задачу, характерную для криминалистической экспертизы аудиоданных. Предположим, имеется цифровой аудиофайл с записью информации. В общем случае это может быть речевой файл, мелодия и вообще любая аудиозапись. На экспертизу также предоставляется аппаратура аудиозаписи. Предполагается, что аудиофайл, представленный на экспертизу, записан на этой аппаратуре. Имеется возможность произвольной аудиозаписи файлов на аппаратуре. Ставится задача определения, на данной ли аппаратуре произведена запись файла, представленного на экспертизу.

В рамках рассматриваемого подхода возможна также постановка задачи сравнения характеристик двух произвольных аудиофайлов, с целью определения — на одной и той же или на разной аппаратуре аудиозаписи произведена запись файлов.

Рассмотрим цифровые аудиофайлы, записанные в формате wav, и проанализируем цифровые аудиоданные как временные ряды амплитуды звуковой волны. Нашей задачей в дальнейшем является выделение статистических характеристик тех параметров аудиофайлов, которые можно было бы с определенной вероятностью интерпретировать как характеристики, обусловленные индивидуальными составляющими шумов цифровой аппаратуры аудиозаписи. Исходя из физических предпосылок, полагаем, что выявление подобных характеристик целесообразно проводить на фрагментах пауз речевых сигналов аудиофайлов.

Паузы речи, как дискретные временные ряды амплитуды звуковой волны A_i , можно в простейшем случае представить в виде суммы собственных шумов цифровой аппаратуры аудиозаписи S_i и внешних источников шумов V_i

$$A_i = S_i + V_i, \quad (1)$$

где i — временной отсчет.

Современная аппаратура аудиозаписи имеет в своем составе множество систем автоматической корректировки различных параметров аппаратуры. Эти системы имеют свои собственные характеристики. Их периодическая (или постоянная) работа может проявляться в наличии в составляющей S_i на различных временных масштабах самоподобных структур. Это характерно для многих временных рядов. При наличии таких структур появляется возможность выявления закономерностей их распределения в этой составляющей, например, различными методами спектрального анализа.

При наиболее общем подходе эти закономерности, с точки зрения характеристик временного ряда пауз, могут рассматриваться как самоподобные мультифрактальные структуры. Будем рассматривать мультифракталы в наших задачах в соответствии с концепцией, введенной впервые Мандельбродом [10–19]. При этом подобие при дальнейшем анализе рассматривается как подобие повторяющихся во времени геометрических структур на различных масштабных уровнях.

Естественно, что составляющие внешних источников шумов V_i , также могут обладать характеристиками мультифрактальности. Однако в разных аудиофайлах характеристики этой составляющей шумов должны различаться между собой независимо от типа аппаратуры аудиозаписи. Но можно предположить, что мультифрактальные характеристики аппаратурной составляющей шумов разных аудиофайлов, записанных на одной и той же аппаратуре аудиозаписи, будут статистически идентичны.

Развиваемая в дальнейшем концепция основывается на сравнении статистических характеристик шумов пауз двух разных аудиофайлов и оценке близости этих характеристик в определенном диапазоне параметров. Решение об идентичности или отличии аппаратуры аудиозаписи двух рассматриваемых файлов принимается с определенной вероятностью ошибки, в зависимости от меры близости этих характеристик.

Постановка задачи состоит в разработке математической модели, позволяющей выявить статистические характеристики шумов аппаратуры двух разных аудиофайлов и принять решение об идентичности их характеристик. Идентичность аппаратуры аудиозаписи является следствием идентичности характеристик шумов в паузах аудиозаписи.

Сегментация аудиофайла на паузы и фрагменты речи

Существенной составляющей частью задачи автоматической идентификации аппаратуры аудиозаписи является сегментация речевых аудиофайлов на паузы и фрагменты речи. От решения этой задачи зависит эффективность решения многих задач в области речевых технологий. Для автоматической сегментации пауз речи будем использовать параметр фрактальной размерности фрагментов аудиофайла [16, 19]. Как будет показано далее, этот параметр существенно отличается для фрагментов речи и пауз.

Далее используем подход оценки фрактальной размерности на основе размерности Хаусдорфа [16, 19]. Введем аппроксимацию графиков временных рядов амплитуды звуковой волны аудиофайлов комплексами, состоящими из двумер-

ных прямоугольников, с геометрическим фактором P . Как следует из определения Хаусдорфа, размерность D определяется из закона

$$R(p) \sim p^{2-D} \text{ при } p \rightarrow 0, \quad (2)$$

где $R(p)$ — полная площадь комплекса с масштабом разбиения p .

На практике, при попытке вычислить D по формуле (2), возникает проблема. Реальные временные ряды всегда имеют минимальный масштаб структуры p_0 . Кроме того, приближение к асимптотическому режиму в (2) обычно является слишком медленным.

Рассмотрим аудиофайл с частотой дискретизации F_s и разрядностью r бит. На практике, эта частота F_s лежит в диапазоне от 8000 Гц до 44100 Гц (иногда выше) при разрядности квантования от 8 до 24 бит.

Очевидно, что минимальная клетка (ячейка) для покрытия графика звуковой волны должна представляться в виде прямоугольника с размерами $a \times b$, где a — минимально возможное, в соответствии с разрядностью оцифровки аудиофайла, изменение амплитуды звуковой волны. Так, например, для 8-разрядного файла $a = 2 / 256$. Здесь, как и далее, рассматривается формат данных wav, в котором амплитуда звуковой волны представлена в виде числа с плавающей запятой в диапазоне от -1 до $+1$. В общем случае, при разрядности аудиофайла r , размер величины

$$a = 2 / 2^r. \quad (3)$$

Минимальный размер прямоугольника по оси времени b равен $1 / F_s$. Поскольку размерность в данном случае не играет роли, примем минимальный размер по оси временного отсчета $b = 1$.

Скорость сходимости аппроксимаций существенно зависит от вида ячеек для аппроксимации. Примем следующий алгоритм расчета фрактальной размерности по Хаусдорфу.

Будем разбивать временное окно с фрагментом паузы аудиофайла на совокупность прямоугольных ячеек со сторонами $a \times b$. Примем масштаб разбиения

$$p = kb, \quad (4)$$

где $k = 1, 2, 3, \dots$ — коэффициент, принимающий целые значения при определении фрактальной размерности.

Определим фрактальную размерность по Хаусдорфу как

$$D = 2 - \lim_{p \rightarrow 0} [\ln(N(p)) / \ln(p)], \quad (5)$$

где $N(p)$ — количество ячеек при покрытии ячейками с масштабом p , которые включают хотя бы одно значение амплитуды звуковой волны; p — масштаб разбиения.

Будем изменять масштаб для конкретной реализации временного окна паузы аудиофайла. Построим график зависимости $\ln(N(p)) = f(\ln(p))$. После построения графика осуществляем аппроксимацию первых 20-ти точек графика линейной зависимостью

$$F(p) = c \ln(p) + c_0, \quad (6)$$

где c и c_0 — коэффициенты аппроксимации.

Количество, равное 20-ти точкам, выбрано экспериментально на основе большого объема исследований. До этого количества точек графики зависимостей носят практически линейный характер.

Величина фрактальной размерности D в рамках данного подхода равна:

$$D = 2 - abs(c), \quad (7)$$

где $abs(c)$ — абсолютная величина коэффициента c .

По приведенной методике было проведено исследование фрагментов речи и пауз аудиофайлов для нескольких сотен различных файлов. Динамика изменения фрактальной размерности фрагментов аудиофайла приведена на рис. 1 и рис. 2 (D — среднее значение фрактальной размерности для всего временного окна, SDI — среднеквадратическое отклонение фрактальной размерности по 6-ти временным подинтервалам временного окна).

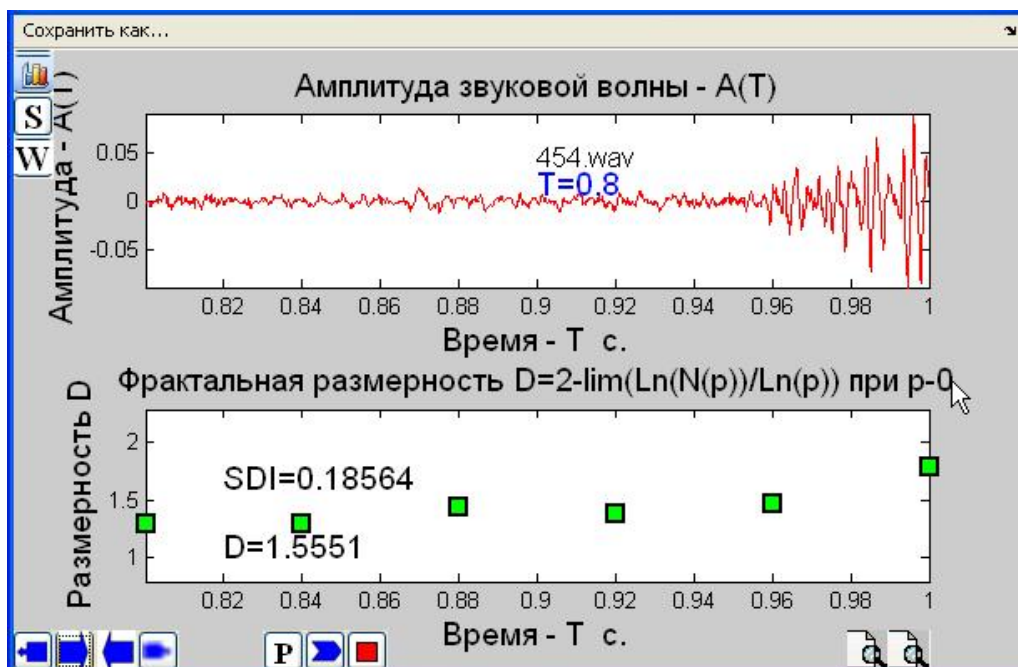


Рис. 1. Иллюстрация динамики изменения фрактальной размерности паузы

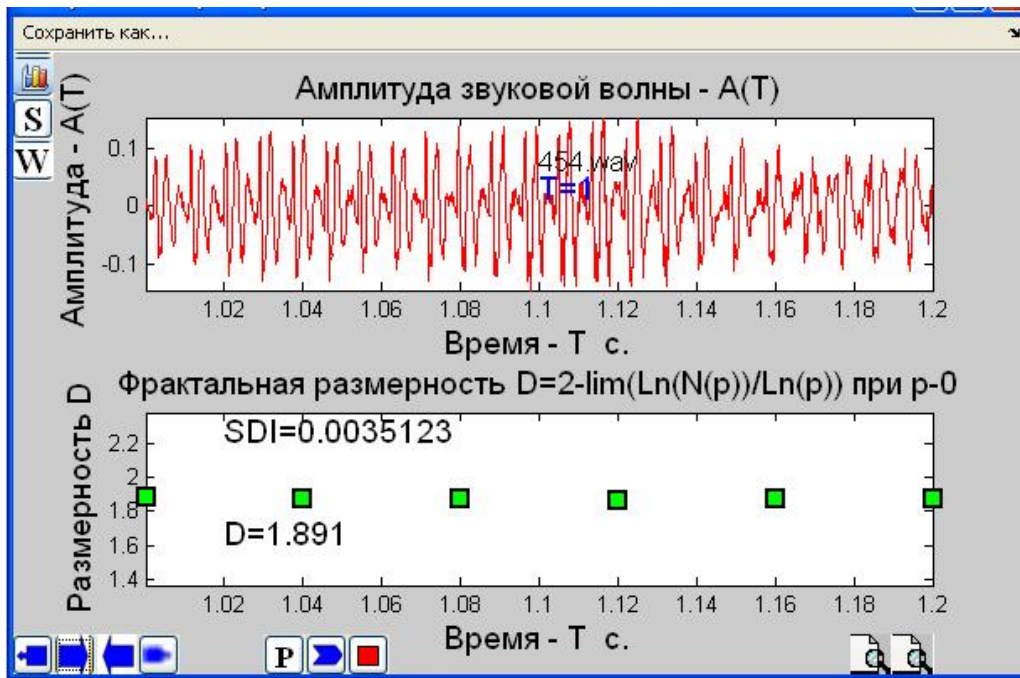


Рис. 2. Иллюстрация динамики изменения фрактальной размерности речевого фрагмента

Анализ результатов исследования показал следующее. Для фрагментов речи аудиофайлов длительностью (20–60) мс фрактальная размерность D находится в среднем в диапазоне от 1,75 до 1,94. На соседствующих временных интервалах длительностью порядка 20 мс ее величина в пределах фрагментов речи, которые можно отождествить с фонемными составляющими, весьма стабильна и изменяется в среднем не более чем на 0,02. При переходе от фрагмента речи к паузе фрактальная размерность резко изменяется. На участках пауз фрактальная размерность D не превышает величины 1,65. Проведенные исследования указывают на возможность эффективной сегментации пауз и фрагментов речи в речевых аудиофайлах по параметру фрактальной размерности.

При исследованиях мультифрактальности пауз применялась автоматическая сегментация аудиофайлов на паузы и фрагменты речи по фрактальной размерности. Временное окно для сегментации варьировалось от 5 до 20 мс в зависимости от частоты дискретизации аудиофайла.

Необходимо отметить следующее. Определение границы паузы и фрагмента речи по любому критерию приносит определенный субъективизм. В данном исследовании границей паузы и фрагмента речи является момент времени, разделяющий два временных окна длительностью (5–20) мс, для которых величина D переходит границу $D_q = 1,65$.

Модель выявления самоподобных структур на основе максимумов вейвлет-преобразования

Рассмотрим совокупность пауз речевого файла, которые обрабатываются на основе модели, изложенной в предыдущем разделе. Будем рассматривать всю со-

вокупность выделенных пауз как дискретный временной ряд амплитуды звуковой волны. Поставим задачу выделения характеристик самоподобных структур в полученном временном ряде. Это могут быть самые различные геометрически подобные структуры, используемые при рассмотрении графиков изменения амплитуды звуковой волны. В соответствии с развиваемой концепцией рассмотрим самоподобие как геометрическое подобие, связанное с преобразованиями сжатия, растяжения как по оси времени, так и по амплитудной координате. Для выявления подобных структур используем методы вейвлет-анализа [20]. С этой целью выберем вейвлет-базис, имеющий существенное сродство с геометрическими структурами амплитуды звуковой волны. Из всего многообразия базисов, используемых на сегодняшний день в практике вейвлет-анализа, по геометрическим структурам к динамике амплитуды звуковой волны в аудиофайлах наиболее близок комплексный вейвлет Морле [20]

$$C_{mor}(t_i, F_b, F_c) = (\pi * F_b)^{0.5} * \exp(2 * j * \pi * F_c * t_i) * \exp((-t_i)^2 / F_b), \quad (8)$$

где F_b — параметр ширины вейвлета; F_c — центральная частота вейвлета; t_i — дискретные временные отсчеты.

Пусть $A(t_i)$ — значение амплитуды звуковой волны фрагмента паузы аудиофайла в момент времени t_i . Рассмотрим одну из модификаций дискретного вейвлет-преобразования для временного ряда пауз на основе комплексного вейвлета Морле

$$C(F_b, F_c) = (1 / N) * \sum_{t_i=0}^N abs \left(\sum_{t_j=0}^{N_m} C_{mor}(t_i - t_j, F_b, F_c) * A(t_i - t_j) \right), \quad (9)$$

где $C(F_b, F_c)$ — среднее значение модуля коэффициента вейвлет-преобразования при фиксированных параметрах F_b, F_c ; N — количество дискретных отсчетов совокупности пауз речевого файла по которым осуществляется усреднение; N_m — количество отсчетов для вычисления свертки вейвлета Морле и дискретных отсчетов амплитуды звуковой волны; t_i, t_j — временные дискретные отсчеты $t_i > t_j$.

На рис. 3 представлен характерный график, иллюстрирующий зависимость $C(F_b, F_c)$ для пауз аудиофайлов.

Основой разрабатываемой модели идентификации аппаратуры аудиозаписи является сравнительный анализ положений локальных максимумов усредненных скейлограмм для двух и более аудиофайлов. Рассмотрим две выборки пауз из двух аудиофайлов. Построим по изложенной выше методике двумерные скейлограммы для функций $C_1(F_b, F_c)$ и $C_2(F_b, F_c)$. Определим совокупность локальных максимумов для каждой из функций — $C_{max1}(F_b, F_c)$, $C_{max2}(F_b, F_c)$ по F_b и F_c . Введем меру близости L_r между положением двух локальных максимумов функций $C_{max1}(F_b, F_c)$ и $C_{max2}(F_b, F_c)$

$$L_r = (F_{b1} - F_{b2})^2 + \delta(F_{c1} - F_{c2})^2, \quad (10)$$

где индексы 1 и 2 относятся соответственно к первой и второй функциям; δ — нормировочный коэффициент, приводящий параметры F_b и F_c к одному порядку по величине.

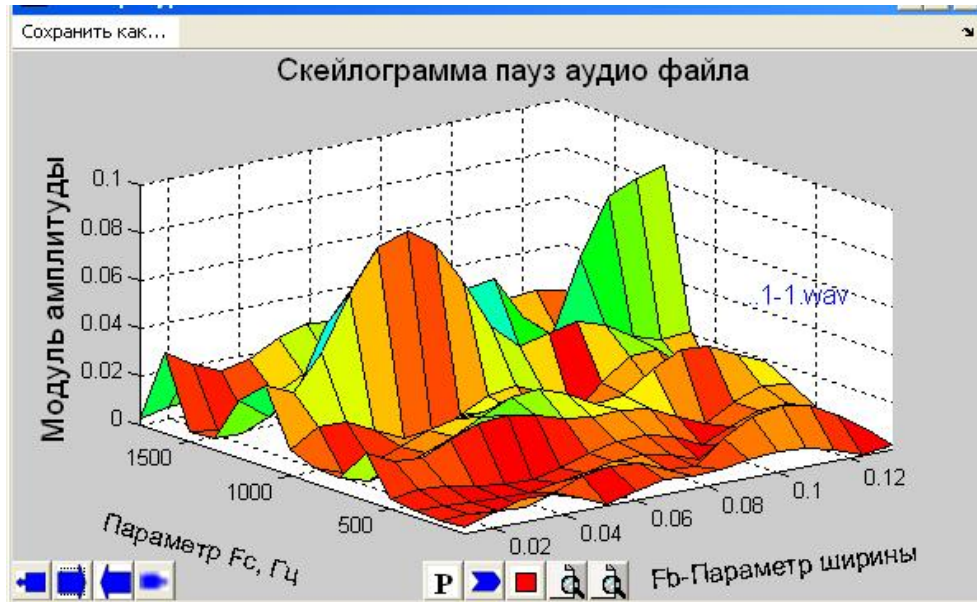


Рис. 3. Скейлограмма пауз аудиофайла

Коэффициент δ определяется экспериментально. Его введение обусловлено необходимостью уравнивания влияния на меру близости экстремумов двух разнородных параметров вейвлета Морле.

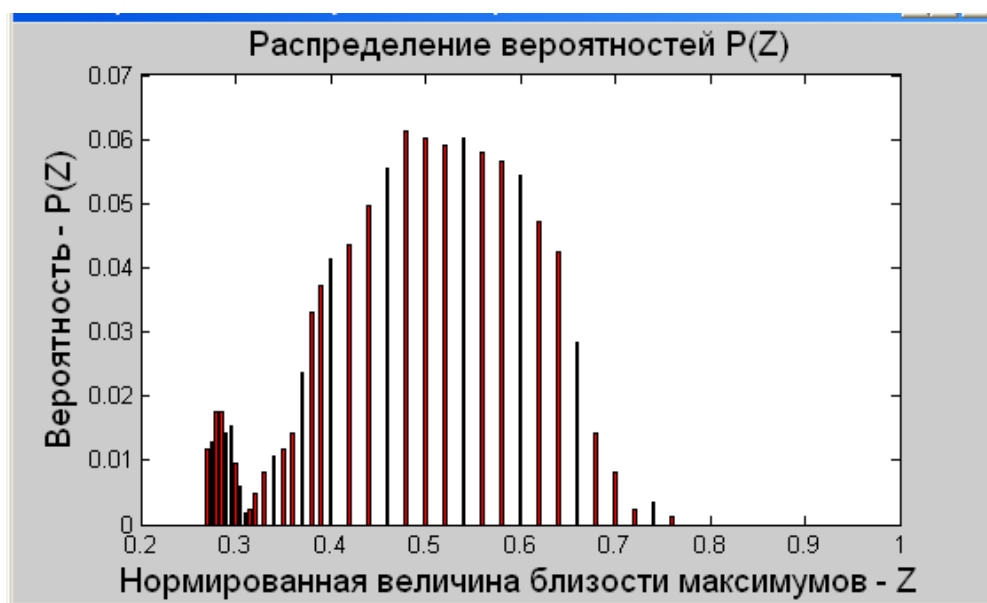
Выберем множество положений локальных максимумов функции $C_{\max 1}(F_b, F_c)$ в качестве базовых с дискретными значениями F_b , F_c и определим дискретное множество $Z(F_{b1}, F_{c1})$ как

$$Z(F_{b1}, F_{c1}) = \inf_{(F_{b2}, F_{c2})} [L_r(F_{b1}, F_{b2}, F_{c1}, F_{c2})]. \quad (11)$$

В рассматриваемой модели $Z(F_{b1}, F_{c1})$ является множеством минимальных расстояний для каждого из локальных максимумов функции $C_{\max 1}(F_b, F_c)$ по отношению к ближайшему локальному максимуму функции $C_{\max 2}(F_b, F_c)$.

Построим функцию распределения вероятностей $P(Z)$ по величинам множества $Z(F_{b1}, F_{c1})$.

На рис. 4 и 5 приведены характерные графики зависимости $P(Z)$ для двух аудиофайлов, записанных на одной и той же аппаратуре и на разной аппаратуре аудиозаписи при различных фоновых условиях.



Многочисленные экспериментальные исследования статистических характеристик пауз различных аудиофайлов при различных условиях записи показали устойчивую особенность распределения вероятностей величины Z . В частности, при записи аудиофайлов на одной и той же аппаратуре аудиозаписи на графике зависимости $P(Z)$ в 86 % присутствует характерный аномальный массив вероятностей в хвостах распределения (в начале графика — рис. 4). Для аудиофайлов, записанных на различной аппаратуре аудиозаписи, этот аномальный, явно выраженный массив в хвостах распределения, отсутствует в 97 % экспериментов (см. рис. 5).

Полученная устойчивая закономерность может являться основой для идентификации аппаратуры аудиозаписи по статистическим характеристикам пауз аудиофайлов.

В рамках разработанной модели для принятия решения о том, на одной и той же или на различной аппаратуре аудиозаписи получены два (или более) аудиофайла, реализуется модель проверки статистических гипотез на основе критерия Колмогорова–Смирнова. С этой целью на графике $P(Z)$, полученном для двух аудиофайлов, записанных на одной аппаратуре аудиозаписи, выделяется фрагмент хвостов распределения $P_x(Z)$ с аномальным массивом в начале распределения. Выделенный участок $P_x(Z)$ перенормируется и рассматривается как распределение вероятностей.

Для проверки аудиофайла, присланного на экспертизу, строим аналогичную функцию $P_{x1}(Z)$ на основе файла, записанного на конкретной аппаратуре аудиозаписи и экспериментального (образцового) файла экспертизы. Эта функция строится в диапазоне изменений аргумента Z функции $P_x(Z)$. Два распределения вероятностей $P_x(Z)$ и $P_{x1}(Z)$ сравниваются при заданной ошибке первого рода по критерию Колмогорова–Смирнова. При подтверждении гипотезы об идентичности распределений с заданной вероятностью ошибки принимается решение о том, что представленный на экспертизу аудиофайл записан на конкретной аппаратуре аудиозаписи. В противном случае гипотеза отвергается.

В качестве иллюстрации на рис. 6 приведен графический фрагмент принятия решения на основе разработанной модели в рамках программной системы идентификации аппаратуры аудиозаписи «Фрактал» [11].

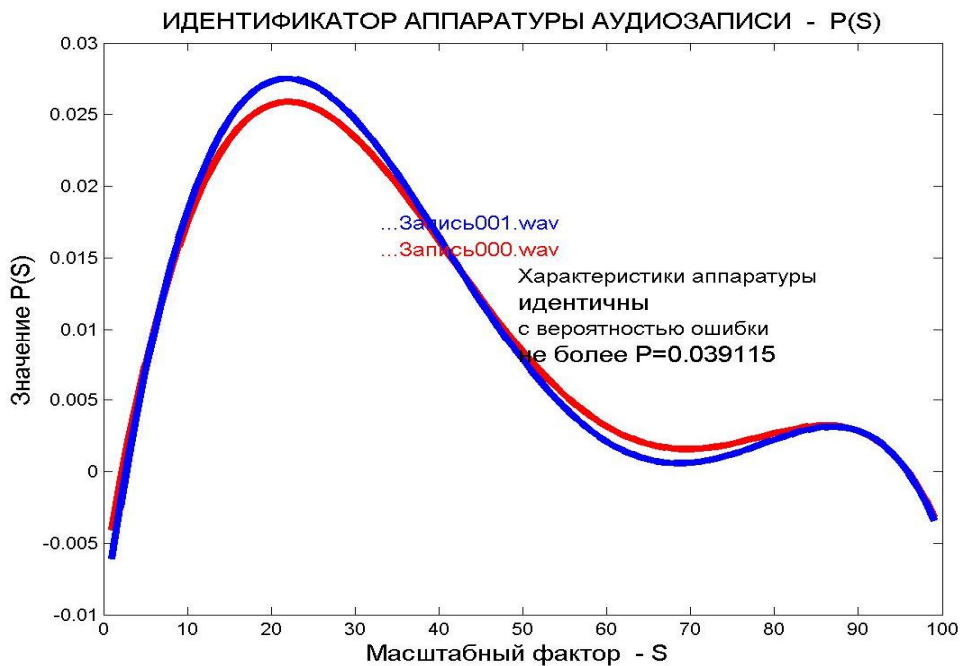


Рис. 6. Иллюстрация идентификации цифровой аппаратуры аудиозаписи в программной системе «Фрактал»

На рис. 6 функция $P(s)$ аналог функции $P_x(Z)$, масштабный фактор S — аналог Z .

Выводы

Разработаны методология и математическая модель выявления и анализа статистических структур в паузах речевых файлов, которые являются следствием индивидуальных характеристик паразитных параметров цифровой аппаратуры аудиозаписи. Проведены массовые экспериментальные исследования, показывающие устойчивость подобных структур в аудиофайлах для файлов, записанных на одной и той же цифровой аппаратуре аудиозаписи. Модель сравнения статистических характеристик этих структур позволяет решать задачи построения систем автоматической идентификации аппаратуры аудиозаписи по статистическим характеристикам сигналов шумов пауз аудиофайлов.

1. Рыбальский О.В. Система идентификации аппаратуры аудиозаписи на основе мультифрактального подхода / О.В. Рыбальский, В.И. Соловьев // Вісник Східноукр. нац. ун-ту. ім. В.Даля. — 2010. — Ч. 1. — № 9(151). — С. 58–63.

2. Рыбальский О.В. Современные методы проверки аутентичности магнитных фонограмм в судебно-акустической экспертизе / О.В. Рыбальский, Ю.Ф. Жариков. — К.: НАВСУ, 2003. — 300 с.

3. Рыбальський О.В. Застосування вейвлет-аналізу для виявлення слідів цифрової обробки аналогових і цифрових фонограм у судово-акустичній експертизі / О.В. Рыбальський. — К.: НАВСУ, 2004. — 167 с.

4. Рыбальський О.В. Методика застосування програми «Академія» («Academy») для виявлення слідів цифрової обробки аналогових і цифрових фонограм при проведенні експертизи матеріалів та засобів відеозвукозапису. Науково-методичні рекомендації експертам-фоноскопістам / О.В. Рыбальський. — К.: КНУВС, 2007. — 40 с.

5. Рыбальский О.В. Идентификация цифровой аппаратуры звукозаписи при использовании программы выявления следов цифровой обработки фонограмм «Академия» / О.В. Рыбальский // Сучасна спеціальна техніка. — 2009. — № 2(17). — С. 92–99.

6. Рыбальский О.В. Применение системного анализа для разработки методик и средств экспертизы технических объектов / О.В. Рыбальский, В.И. Соловьев, Т.А. Татарникова // Теорія та практика судової експертизи і криміналістики: зб. наук. праць. — Вип. 11. — Х.: Право, 2011. — С. 348–353.

7. Методология проверки подлинности сигналограмм выделением самоподобных структур / О.В. Рыбальский, Е.В. Белозеров, В.И. Соловьев, Я.А. Белозерова // Захист інформації. — 2010. — № 2. — С. 35–43.

8. Рыбальский О.В. Организация и методология проведения исследований пригодности программы «Фрактал» для идентификации цифровой аппаратуры записи / О.В. Рыбальский // Сучасний захист інформації. — 2010. — Спецвипуск (4). — С. 118–122.

9. Система идентификации аппаратуры записи сигналов и проверки оригинальности и подлинности сигналограмм / О.В. Рыбальский, В.И. Соловьев, Т.В. Командина, Т.А. Татарникова // Інформаційна безпека. — 2011. — № 2(6). — С. 7–14.

10. Соловьев В.И. Сегментация звукового сигнала в задачах выявления монтажа в аудиофайлах / В.И. Соловьев, О.В. Рыбальский // Вісник Київського нац. ун-ту ім. Т. Шевченко. Військово-спеціальні науки. — 2011. — № 33. — С. 213–220.
11. Надежность идентификационных исследований аппаратуры звукозаписи с применением программы «Фрактал» / О.В. Рыбальский, В.И. Соловьев, А.Н. Шабля [и др.] // Вісник Східноукр. нац. унів. ім. В. Даля. — 2012. — Ч. 1. — № 8(179). — С. 79–84.
12. Сегментация речи в задачах выявления монтажа аудиофайлов / В.И. Соловьев, О.В. Рыбальский, В.В. Журавель, Т.В. Командина // Інформаційна безпека. — 2012. — № 1 (7). — С. 35–42.
13. Рыбальский О.В. К развитию теории, методов и средств проведения экспертизы материалов цифровых фото, видео и звукозаписи / О.В. Рыбальский, В.И. Соловьев // Сучасна спеціальна техніка. — 2012. — № 3(30). — С. 119–121.
14. Федер Е. Фракталы / Е. Федер. — М.: Мир, 1991. — 326 с.
15. Mandelbrot B. Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis / B. Mandelbrot. — Annals of Economic Social Measurement 1, 1972.
16. Mandelbrot B. The Fractal Geometry of Nature / B. Mandelbrot. — New York: W.H. Freeman, 1982.
17. Mandelbrot B. A Multifractal Walk Down Wall Street / B. Mandelbrot. — Scientific American, 1999.
18. Mandelbrot B.B. Robustness of the Rescaled Range R/S in the Measurement of Non-Cycling Long-Run Statistical Dependence / B.B. Mandelbrot // Water Resources Research, 1969. — N 5. — P. 967–988.
19. Павлов А.Н. Мультифрактальный анализ сложных сигналов / А.Н. Павлов, В.С. Анищенко // Успехи физических наук. — 2007. — Т. 177, № 8.
20. Малла С. Вейвлеты в обработке сигналов / С. Малла. — М.: Мир, 2005. — 670 с.

Поступила в редакцию 27.12.2012