

ИНФОРМАТИКА

УДК 684.3

ЛОКАЛИЗАЦИЯ СЛЕДОВ ОБРАБОТКИ СИГНАЛА В ЗАДАЧАХ МОНТАЖА АУДИОЗАПИСИ

*канд. техн. наук, доц. В.И. СОЛОВЬЕВ**(Восточноукраинский национальный университет имени В. Даля, Луганск);**д-р техн. наук, проф. О.В. РЫБАЛЬСКИЙ**(Национальная академия внутренних дел, Киев);**д-р техн. наук, проф. В.К. ЖЕЛЕЗНЯК**(Полоцкий государственный университет)*

Приведены результаты исследований по локализации следов цифровой обработки сигналов в задачах монтажа аудиофайлов. Направление исследований основывается на комплексном подходе при анализе и локализации монтажа. Подход базируется на различии статистических характеристик участков паузы между словами и речью.

В последние годы серьезное внимание уделяется вопросам разработки методов и средств автоматизации задачи выявления монтажа аудиозаписи [1]. Существует множество исследований и подходов к решению этой задачи. Однако на сегодняшний день неясно, в какой степени разрешима эта проблема и при каких условиях возможно обнаружение и локализация монтажа аудиосигнала.

Рассмотрим классическую постановку задачи. В звуковом файле с записью речевого сигнала осуществлена вставка фрагментов записи из другого звукового файла или осуществлена перестановка фрагментов записи из различных частей одного и того же файла. Требуется методами цифровой обработки звукового сигнала локализовать вероятное место монтажа. Далее приводится анализ эффективности различных подходов и излагаются результаты исследования возможностей нового направления.

Спектральные методы локализации монтажа. Возможность эффективного применения спектрального анализа к задаче выявления монтажа аудиофайла имеет специфику, существенно усложняющую решение. Монтаж, реализованный с использованием современных программных продуктов, локализован на малом временном интервале. Этот интервал зависит от частоты дискретизации файла и находится обычно в диапазоне от 0,0001 до 0,001 с для частоты дискретизации 44,1 кГц и 0,05...0,1 с для частоты 8 кГц. По порядку величины временного интервала это соизмеримо с $1/F$, где F – частота дискретизации.

На рисунках 1 и 2 приведены примеры «необработанного» монтажа двух фрагментов звукового файла.



Рис. 1. Иллюстрация монтажа в паузе (место монтажа – $T = 0,0625$ с)

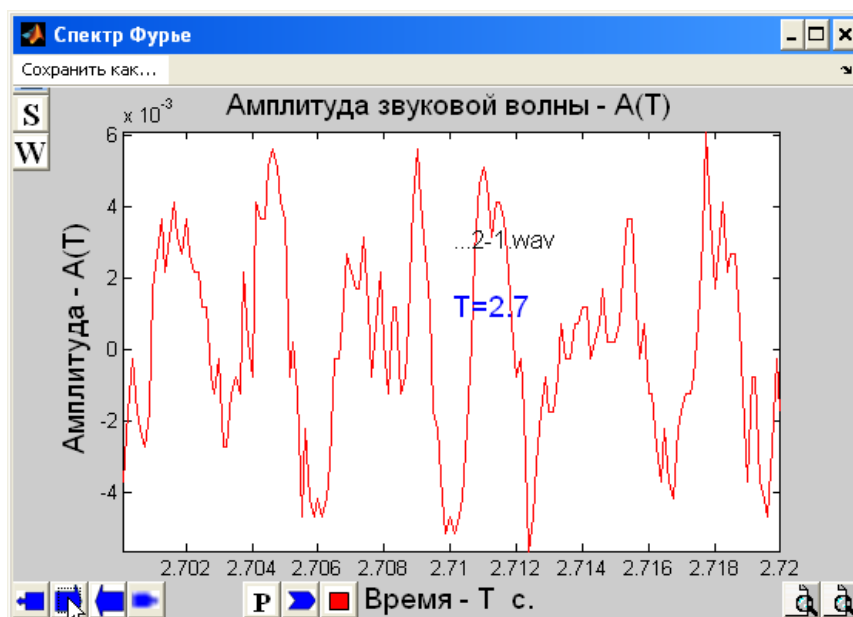


Рис. 2. Иллюстрация монтажа в паузе (место монтажа – $T = 2,7120$ с)

В первом случае (см. рис. 1) визуально видно изменение характеристик аудиосигнала. Во втором (см. рис. 2) – практически никаких визуальных отличий нет. Здесь были подобраны фрагменты речи с близкими визуальными характеристиками пауз. При прослушивании второго фрагмента наличие вероятного монтажа выявляется ввиду логического несоответствия последовательности слов речи. Задача состояла в разработке автоматизированных методов и средств локализации мест вероятного монтажа. Поэтому данная иллюстрация показательна. Простая вставка визуально идентичных фрагментов не имеет никаких явных визуальных признаков монтажа. Применение методов спектрального анализа для выявления следов монтажа, неразличимых визуально, иногда может оказаться эффективным и в ряде конкретных случаев дает результат. Однако человеческий глаз обладает весьма высокими характеристиками по выявлению «аномалий», возможных в месте локализации монтажа. Маловероятно, что при отсутствии явно выраженных визуальных «аномалий», обусловленных наличием монтажа, методы спектрального анализа выявят его предполагаемое место в аудиозаписи. В большинстве случаев разрешающая способность спектральных методов не обеспечивает эффективное решение этой задачи. Эти замечания носят эвристический характер, однако, как будет показано далее, постановка задачи обнаружения места вероятного монтажа является весьма сложной и проблематичной.

Отдельного рассмотрения заслуживают методы обнаружения монтажа, основанные на идее изменения естественных фазовых характеристик сигнала при вставке «инородной» аудиозаписи. Достаточно многочисленные исследования в этой области стремятся доказать эффективность этого подхода. Но большинство исследователей не берут во внимание весьма важный фактор, который фактически ставит барьер на пути применения этих методик. Дело в том, что на сегодняшний день практически любая запись с запоминанием аудиоинформации осуществляется в сжатых форматах аудиофайлов. Наиболее широко используемый формат – MPEG формат аудиофайлов и его современные производные. Этот формат базируется на стандарте сжатия аудиоинформации, основой которого являются нейрофизиологические законы восприятия звуковой информации. Важной особенностью алгоритмов стандарта MPEG (как и других форматов сжатия) является несущественность фазовой части аудиоинформации при восприятии звука человеком (за исключением определения направления на источник звука). Эта информация после частотного преобразования Фурье исключается как несущественная. Таким образом, при исследовании аудиофайла (после обратного преобразования Фурье) искажение естественных закономерностей фазы, обусловленное монтажом аудиозаписи, не может быть обнаружено. После применения сжатия в современных форматах анализ фазы в большинстве задач речевых технологий лишен смысла.

Постановка задачи исследования. На рисунке 3 представлен характерный пример графика речевого аудиосигнала. Анализ подобных сигналов показывает, что в большинстве случаев характеристики сигнала в паузах существенно отличаются, как минимум, по одному параметру – амплитуде сигнала в паузах. Это обусловлено вариабельностью шумов фона и в ряде случаев нестабильностью характеристик аппаратуры аудиозаписи сигнала на интервалах времени нескольких речевых фрагментов. При рассмотрении одной паузы эти характеристики достаточно стационарны в пределах паузы. Данное утверждение является эвристическим в постановочной части задачи исследования. Результаты исследо-

ваний многочисленных аудиофайлов подтверждают эту физическую посылку исследования. Ниже приведены результаты исследований по локализации монтажа аудиозаписи, которые основываются на сравнении статистических характеристик двух равных частей паузы. Исследуемый аудиофайл разбивается на сегменты, состоящие из пауз речи. Алгоритм выделения пауз в задачах автоматического выявления монтажа играет весьма важную роль. Точность разделения границ фрагментов речи и пауз в аудиофайлах оказывает очень сильное влияние на любые характеристики пауз при любом методе исследования.

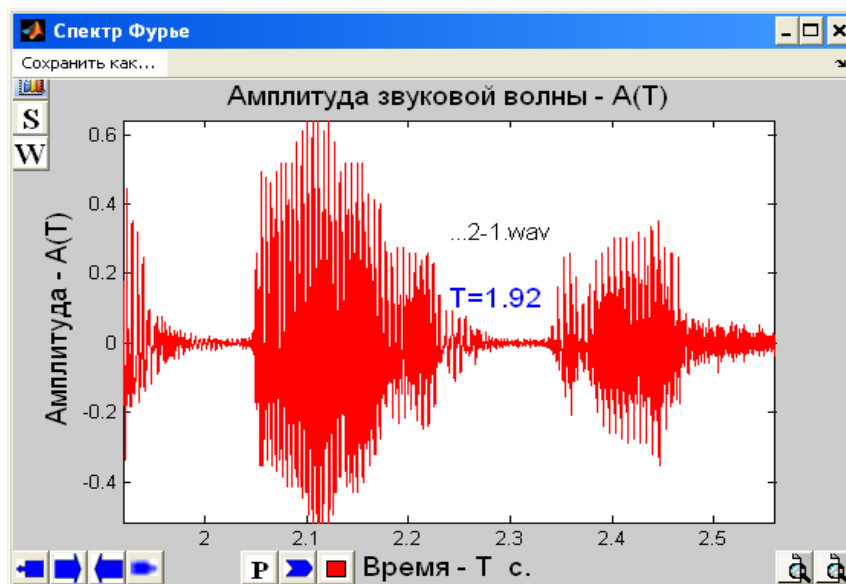


Рис. 3. График амплитуды речевого сигнала

В приводимых ниже исследованиях сегментация аудиофайла на фрагменты речи и паузы основывается на подходе, базирующемся на существенном отличии мультифрактальных характеристик пауз и фрагментов речи. Изложение сущности этой методологии является предметом отдельной работы и в данном исследовании не приводится.

Выявление монтажа осуществляется на основе комплексного подхода с анализом ряда статистических характеристик пауз. В задачах исследования монтажа важно применение методов анализа с высоким разрешением. Далее при оценках статистических характеристик пауз используются методы вейвлет-анализа.

После сегментации аудиофайла на фрагменты, состоящие из пауз, осуществляется разбиение каждой паузы на два равных по времени интервала. На каждом из интервалов рассчитываются две статистические характеристики: 1) среднее значение амплитуды звуковой волны и её дисперсия; 2) среднее значение характеристик деформации скейлограммы на основе вейвлет-преобразования с комплексным базисом Морле и дисперсия этой величины [2 – 4].

После расчета этих характеристик осуществляется сравнение значений соответствующих характеристик для двух половин каждой паузы с целью проверки статистической гипотезы о равенстве средних значений. Соответственно, при неравенстве средних значений хотя бы по одной из характеристик принимается (автоматически) решение о наличии возможного монтажа аудиозаписи. Такова общая концепция разработанного подхода.

Рассмотрим оценки по каждой из статистических характеристик. Первая характеристика, как показывают исследования, чувствительна к изменениям амплитудных характеристик фоновых сигналов при монтаже аудиосигнала. Вторая – выявляет возможные спектральные изменения при монтаже в паузах. Эти две характеристики охватывают большинство вариантов изменения характеристик сигнала при монтаже. Исключение – случаи, когда характеристики практически неизменны, но вследствие монтажа возможна какая-либо локальная аномалия в точке монтажа. На наш взгляд, подобная аномалия при условии выявления вряд ли может быть свидетельством о наличии монтажа.

Необходимо также отметить, что проведение данных исследований было бы невозможно без использования специализированных программных средств автоматизации фоновскопических исследований. Излагаемые результаты получены с применением системы автоматической идентификации аппаратуры аудиозаписи «Фрактал» [5].

Исследование статистических характеристик пауз с признаками монтажа. Анализ аудиоинформации с высоким пространственно-временным разрешением требует применения методов вейвлет-анализа [4]. Важную роль при выборе базиса для вейвлет-преобразования играет степень сродства базиса

с конкретным типом сигналов [4]. Анализ существующих базисов указывает на возможную эффективность в задачах обработки аудиоинформации комплексного вейвлета Морле [4]. В рамках данного исследования была произведена оценка эффективности этого базиса для решения поставленных задач. Общая методология применения данного вейвлет-базиса состоит в следующем.

Рассмотрим введенную в постановочном разделе характеристику – среднее значение деформации скейлограммы на основе вейвлет-преобразования с комплексным базисом Морле.

Преобразование Морле осуществлялось на основе комплексного вейвлета Морле [3]:

$$C_{mor}(t_i, F_b, F_c) = (\pi F_b)^{0.5} \exp(j2\pi F_c t_i) \exp\left(-\frac{t_i^2}{F_b}\right), \quad (1)$$

где F_b – параметр ширины вейвлета; F_c – центральная частота вейвлета; t – дискретные временные отсчеты; j – комплексная единица.

Для каждого временного отсчета паузы на малом временном интервале осуществлялось дискретное преобразование Морле с варьированием в определенном диапазоне параметров F_b и F_c . Строилась для каждого временного отсчета (в каждый момент времени) скейлограмма в виде трехмерной зависимости модуля амплитуды вейвлет-преобразования от параметров F_b и F_c (рис. 4).

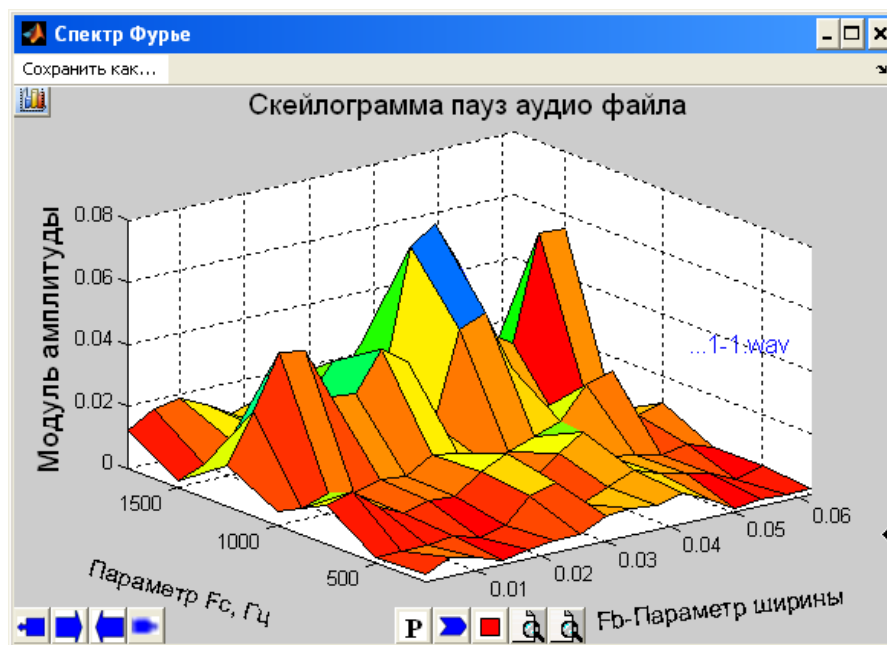


Рис. 4. Скейлограмма фрагмента паузы аудиофайла в координатах F_b, F_c (фиксированный временной отсчет на малом временном интервале)

После этого методами статистического анализа осуществлялось сравнение статистических характеристик распределения подобных скейлограмм для двух половин паузы. И на основе критериев принятия статистических гипотез принималось решение о наличии отличий статистических характеристик скейлограмм двух частей паузы.

Однако подобный строгий подход обладает высокой вычислительной сложностью, которая не может быть реализована при практических вычислениях на современных компьютерах. Причиной сложности является двухпараметричность вейвлета Морле и неортогональность составляющих его базиса. Ввиду этого был разработан вариант с приемлемой вычислительной сложностью. Выбранная половина паузы делилась на несколько равных по времени отрезков. Для каждого из отрезков – временных интервалов выбранной половины паузы – вычислялись дискретные свертки вейвлета Морле с аудиосигналом (1) и в каждой временной точке рассчитывался модуль амплитуды коэффициентов преобразования при фиксированном параметре F_b с дискретным варьированием центральной частоты в диапазоне от 100 до 4000 Гц (для файлов с частотой дискретизации 8 кГц) и от 100 до 20000 Гц (для файлов с частотой дискретизации 40 кГц) с различным шагом дискретизации по частоте. Число разных частот, для которых выполнялись преобразования, равнялось 70. Выбор этой величины обусловлен достаточно высокой вычислительной сложностью, с одной стороны, и достаточным для практических целей приближением, с другой стороны, при построении скейлограммы.

После вычисления модуля амплитуды вейвлет-преобразования с базисом Морле вычислялась величина – «характеристика деформации» скейлограммы:

$$D_m = \frac{\sum_{F_c} A(F_c) \cdot F_c}{\sum_{F_c} A(F_c)}, \quad (2)$$

где D_m – количественная характеристика деформации скейлограммы; $A(F_c)$ – модуль амплитуды вейвлет-преобразования для частоты F_c .

Суммирование осуществляется для всего набора частот.

Предлагаемая характеристика определяет «геометрический центр тяжести скейлограммы» на определенном временном интервале по частоте. При смещении частотных характеристик в ту или иную сторону изменяется и эта характеристика.

После указанных расчетов в каждой временной точке подынтервалов паузы вычислялось среднее значение этой характеристики и ее дисперсия для каждой половины паузы (G , D_{mm}). Эти величины для каждой половины паузы позволяют на основе, например, статистического критерия Стьюдента осуществить проверку гипотезы об идентичности спектральных характеристик сигнала в двух частях паузы.

На рисунках представлены в качестве иллюстрации пространственные скейлограммы для двух участков пауз с наличием монтажа и отличными в двух половинах паузы деформациями спектра (рис. 5) и без монтажа (рис. 6).

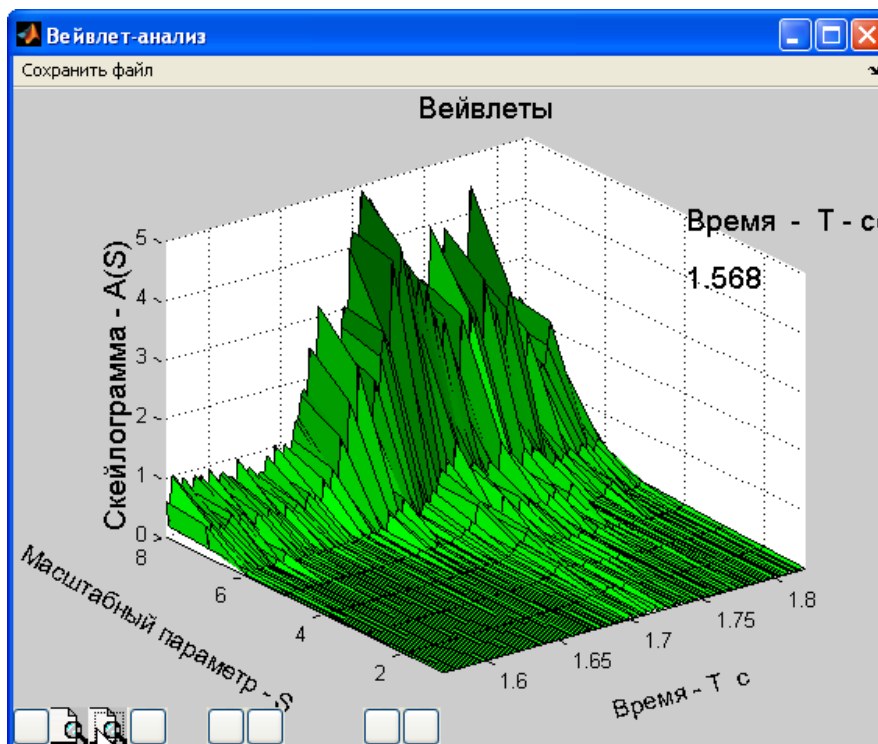


Рис. 5. Пространственная скейлограмма участка паузы с монтажом аудиозаписи

Скейлограммы построены для 8 масштабных факторов, которые на графике пронумерованы в порядке уменьшения эквивалентной центральной частоты вейвлета Морле. На иллюстрациях отчетливо видно отличие в двух половинах паузы характеристики для варианта с монтажом (см. рис. 5). Необходимо подчеркнуть важный момент для рассматриваемого критерия деформации скейлограммы – этот критерий чувствителен к перераспределению энергии между частотами. Определение величин критерия деформации и его дисперсии для первого и второго варианта иллюстраций и проверка гипотезы равенства этой величины для двух половин каждой из пауз по критерию Стьюдента показывают отличие этих величин для первого (см. рис. 5) и их равенство для второго (см. рис. 6) графиков с вероятностью ошибки второго и первого рода соответственно не более 0,023. Приведенный критерий, связанный с частотными характеристиками аудиосигнала, в некоторых случаях, как показывают исследования, уступает по эффективности критерию равенства средних значений амплитуд в двух половинах паузы. Так, например, на рисунке

7 представлена иллюстрация, на которой реализован монтаж из двух участков аудиофайла с практически идентичными частотными характеристиками, но с различным по интенсивности уровнем шума. Критерий деформации скейлограммы показывает идентичность величин деформации в двух частях паузы, но средние значения звуковой амплитуды в двух половинах паузы существенно отличаются. В этом случае, достаточно часто встречающемся на практике, необходимо применять среднее значение абсолютной величины амплитуды звуковой волны (и дисперсию этой величины).

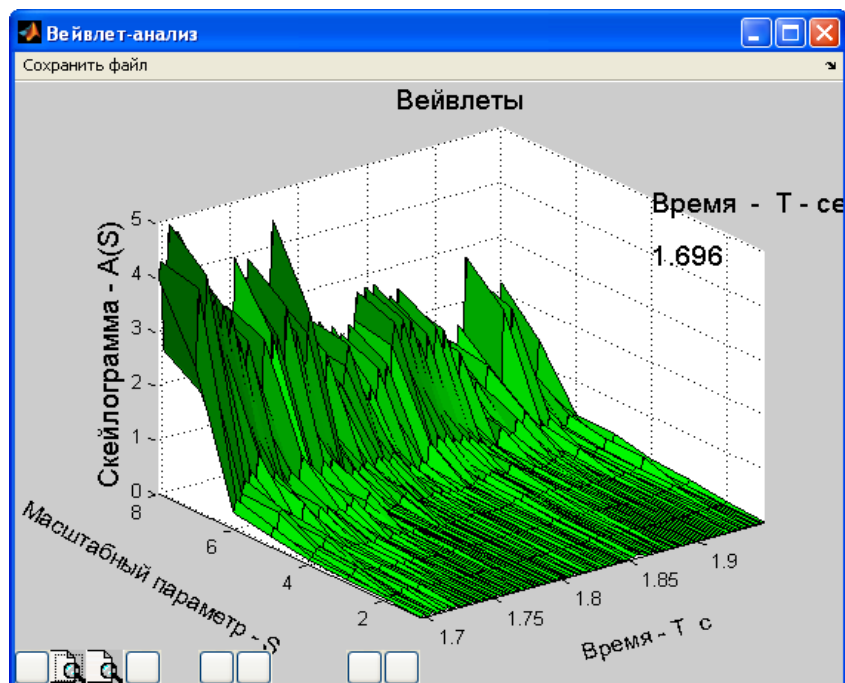


Рис. 6. Пространственная скейлограмма участка паузы без монтажа аудиозаписи

На рисунках 7 и 8 продемонстрирована работа «идентификатора монтажа» на участке паузы с монтажом аудиозаписи в рамках системы идентификации аппаратуры аудиозаписи «Фрактал» [5].

На рисунке 7 показаны распределения «идентификатора монтажа» для паузы без монтажа; на рисунке 8 – с монтажом. В качестве идентификатора монтажа выступает распределение среднего значения амплитуд звуковой волны в двух половинах паузы.

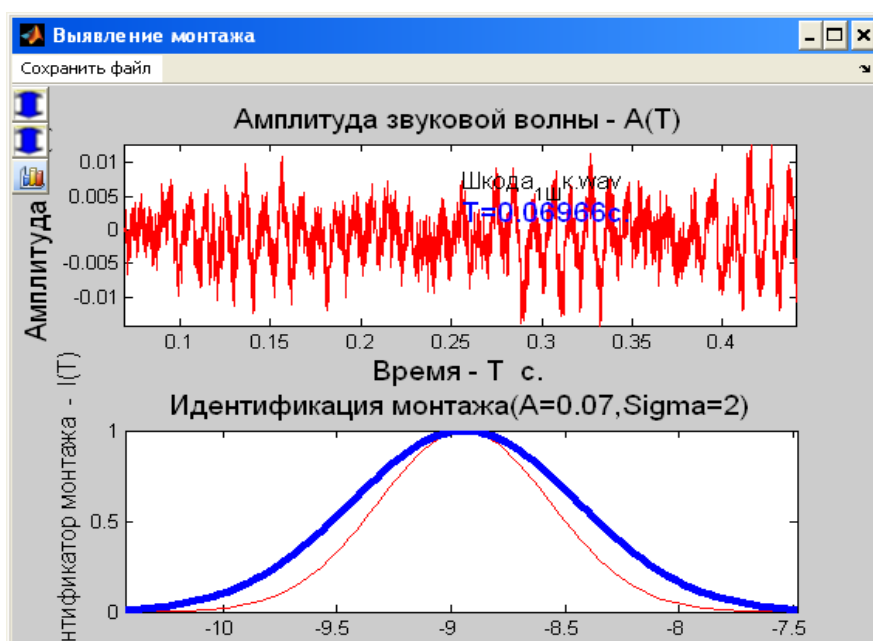


Рис. 7. Распределение среднего значения амплитуды звуковой волны (пауза без монтажа)

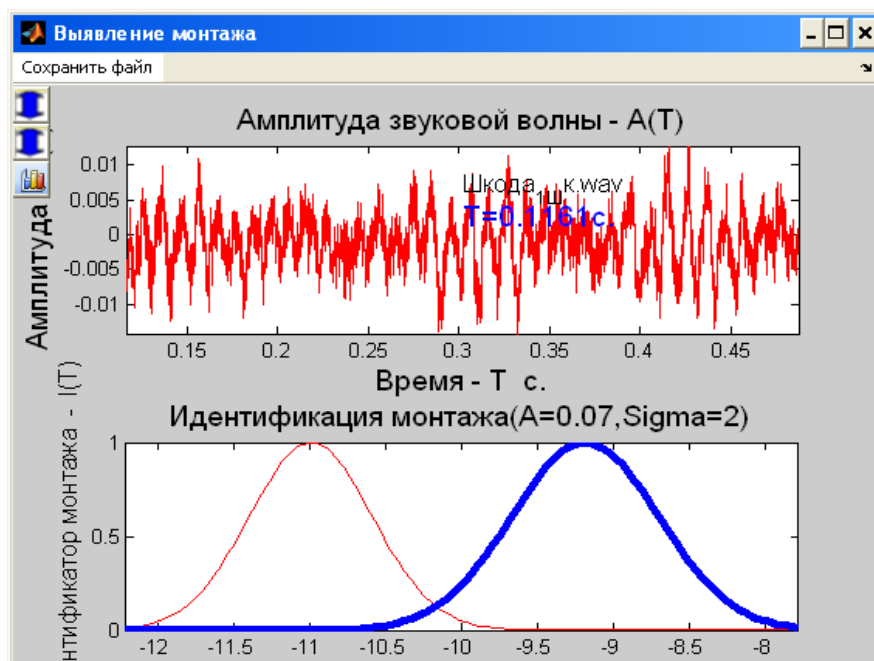


Рис. 8. Распределение среднего значения амплитуды звуковой волны (пауза с монтажом)

Закключение. Решение задачи автоматического выявления и локализации монтажа аудиофайла требует комплексного подхода с применением различного набора статистических характеристик пауз. Необходим анализ изменения как частотных характеристик аудиосигнала в пределах паузы, так и амплитудных изменений, которые могут не учитываться частотными характеристиками. Применение методик выявления монтажа по статистическим характеристикам различных пауз, как косвенно указывает данное исследование, малоэффективно. Статистические характеристики пауз обладают значительной вариабельностью в пределах одного и того же аудиофайла, применительно к задачам выявления монтажа (в других задачах анализа, например различных аудиофайлов, это может оказаться эффективным). Для эффективного решения задач локализации мест вероятного монтажа аудиофайла целесообразно применение методов вейвлет-анализа с высоким разрешением как по частоте, так и по времени.

ЛИТЕРАТУРА

1. Кобозева, А.А. Разработка общей теории выявления следов цифровой обработки сигналограмм и ее реализация аппаратно-программным комплексом «Теорема-М» / А.А. Кобозева, О.В. Рыбальский, В.И. Соловьев // Сучасна спеціальна техніка. – Киев, 2010. – № 1 (20). – С. 5 – 14.
2. Основы цифровой обработки сигналов / А.И. Солонина [и др.]. – 2-е изд. – СПб.: БХВ-Петербург, 2005. – 386 с.
3. Айфичер, Э. Цифровая обработка сигналов / Э. Айфичер, Б. Джервис. – СПб.: БХВ-Петербург, 2004. – 989 с.
4. Мала, С. Вейвлеты в обработке сигналов / С. Мала. – М.: Мир, 2005. – 671 с.
5. Рыбальский О.В. Система идентификации аппаратуры аудиозаписи на основе мультифрактального подхода / О.В. Рыбальский, В.И. Соловьев // Вісник Східноукраїнського національного університету. – 2010. – № 9 (151). – С. 58 – 64.

Поступила 21.01.2013

LOCALIZATION OF TRACKS OF SIGNAL PROCESSING IN TASKS OF ARRANGEMENT OF AUDIO RECORDING

V. SOLOVYOV, O. RYBALSKY, V. ZHELEZNYAK

The results of studies of the localization of the tracks of digital signal processing in tasks of arrangement of audiofiles are shown. The direction of research is based on an integrated approach for the analysis and localization of the arrangement. The approach is based on the difference of the statistical characteristics of parts of speech pauses between words and speech.