УДК 621.317.799 297 + 681.849

О.В. Рыбальский, доктор технических наук, профессор, профессор Национальной академии внутренних дел, г. Киев,
В.И. Соловьев, кандидат технических наук, доцент, заместитель заведующего кафедрой Восточноукраинского национального университета им. В. Даля, г. Северодонецк,
В.В. Журавель, кандидат технических наук, заместитель заведующего лабораторией ГНИЭКЦ МВД Украины, г. Киев

АВТОМАТИЧЕСКАЯ СЕГМЕНТАЦИЯ ФОНОГРАММ ПО ПАУЗАМ РЕЧЕВОГО ПОТОКА

Рассмотрены вопросы создания системы автоматического выделения пауз из речевого потока, записанного на фонограмме. Показано, что использование автоматической сегментации пауз является необходимым условием создания современных средств экспертизы материалов и аппаратуры аналоговой и цифровой звукозаписи. Рассмотрены характеристики пауз, образующихся в свободном осмысленном речевом потоке, и показано, что минимальная длительность паузы, подлежащей выделению, составляет 50 мсек. Определены характеристики сигналов пауз, которые могут использоваться при построении системы автоматической сегментации. Показаны особенности сигналов, содержащихся в паузах фонограмм. Представлены разработанные методы и средства, использованные при создании такой системы, и проиллюстрированы результаты ее работы.

Ключевые слова: *автоматическая сегментация, аппаратура звукозаписи, пауза, фонограмма, сигнал речи, сигнал паузы, экспертиза.*

Розглянуто питання створення системи автоматичного виділення пауз з мовленнєвого потоку, записаного на фонограмі. Показано, що використання автоматичної сегментації пауз є необхідною умовою створення сучасних засобів експертизи матеріалів та апаратури аналогового й цифрового звукозапису. Розглянуто характеристики пауз, що утворюються у вільному осмисленому мовленнєвому потоці, і показано, що мінімальна тривалість паузи, яка підлягає виділенню, складає 50 мсек. Визначено характеристики сигналів пауз, які можуть використовуватися при побудові системи автоматичної сегментації. Показано особливості сигналів, що містяться в паузах фонограм. Представлено розроблені методи та засоби, використані при створенні такої системи, та проілюстровано результати її роботи.

Ключові слова: *автоматична сегментація, апаратура звукозапису, пауза, фонограма, сигнал мовлення, сигнал паузи, експертиза.*

The issues of creation of the system of automatic selection of pauses are considered from a speech stream written on a phonogram. It is shown that the use of automatic segmentation of pauses is the necessary condition of creation of modern facilities of examination of materials and apparatus of the analog and digital audio recording. Descriptions of pauses appearing in a free intelligent speech stream are considered, and it is shown that minimum duration of pause subject to the selection, 50 ms makes.

Descriptions of signals of pauses which can be used for the construction of the system of automatic segmentation are certain. The features of the signals contained in the pauses of phonograms are shown. The worked out methods and facilities, used for creation of such system, are presented, and her job performances are illustrated.

Keywords: *automatic segmentation, apparatus of the audio recording, pause, phonogram, signal of speech, signal of pause, examination.*

Автоматическая сегментация фонограмм по паузам речевого потока является одной из ключевых задач, решение которой необходимо при создании многих систем, предназначенных для проведения экспертизы материалов и средств видео и звукозаписи, например:

– идентификации личности по физическим параметрам сигналов устной речи;

– идентификации аппаратуры записи и проверки оригинальности фонограмм;

– выявления следов монтажа фонограмм и установления их подлинности.

При ее решении необходимо, во-первых, определить минимальную длительность паузы, подлежащей выделению, и, во-вторых, определить или разработать метод и средство для выделения участков пауз из фонограммы. Экспериментальным путем было установлено, что минимальная длительность паузы, которую можно использовать для монтажа по слогам, составляет 50 мсек. Решить задачу выделения участков пауз из фонограммы пытались разными способами. Прямой способ ее решения – установка фиксированного порога по уровню сигнала в паузах фонограммы с последующей корректировкой этого уровня для каждой конкретной паузы в ручном режиме. Такой метод был использован при создании комплекса "Теорема", используемого для проведения экспертиз материалов и аппаратуры аналоговой магнитной записи (ААМЗ) [1]. Такое решение, хотя и было достаточно трудоемким, но обеспечило выделение пауз с точностью, достаточной для работы этого комплекса.

Однако сама идеология построения программного обеспечения этого комплекса не требовала высокой точности выделения пауз. Она не могла обеспечить возможности проведения экспертных исследований аппаратуры цифровой звукозаписи (АЦЗЗ) и цифровых фонограмм, а также выявления следов монтажа по слогам и монтажа, выполненного путем компиляции записей, сделанных на одной и той же ААМЗ, или из записей, произведенных на АЦЗЗ. Отметим, что минимальная длительность пауз, выделяемых в комплексе "Теорема", составляла 120 мсек (что слишком много для выявления монтажа по слогам).

При разработке новых методов и средств экспертизы материалов и аппаратуры видео и звукозаписи необходимо обеспечить автоматическую сегментацию фонограммы по паузам речи. Эта необходимость обусловлена современными требованиями к системе проведения такой экспертизы:

– обеспечением идентификационных исследований как аналоговой, так и цифровой аппаратуры звукозаписи;

– обеспечением диагностических исследований как аналоговых, так и цифровых фонограмм;

– обеспечением выявления следов монтажа, выполненного любым из известных способов компиляции нового содержимого из одной или различных фонограмм.

Последнее требование подразумевает выявление монтажа, выполненного по слогам, монтажа, выполненного способом вырезания и перестановки фрагментов из монтируемой фонограммы или фонограмм, записанных на одной аппаратуре, и монтажа, выполненного из цифровых фонограмм с последующей перезаписью фальсификата на аналоговую аппаратуру.

Обеспечить эти требования без автоматизации процесса проведения экспертизы невозможно из-за ее огромной трудоемкости, а основным первоначальным условием автоматизации экспертизы является автоматическая сегментация фонограмм по паузам.

Для построения системы сегментации следует более подробно рассмотреть понятие паузы в речевом потоке. Очевидно, что речевая информация и пауза различаются по информационному содержанию, уровням, спектральному и фрактальному составам содержащихся в них сигналов. При этом информационное содержание паузы с точки зрения эксперта несет в себе характеристики окружающей звуковой среды и характеристики аппаратуры, используемой при записи каждой конкретной фонограммы [1]. Таким образом, параметры паузы могут быть описаны ее длительностью, уровнями, спектральным и фрактальным составом содержащихся в ней сигналов.

Из экспертной практики известно, что в осмысленном речевом потоке встречаются два основных вида пауз: паузы между отдельными словами и паузы между слогами. Мы исходим из того, что монтаж фонограммы может осуществляться путем компиляции нового содержимого [2] как из слогов, так и из отдельных слов и предложений. Очевидно, что и в первом, и во втором случае монтаж (т.е. "сшивка" двух фрагментов) может производиться только в паузах. Этим и объясняется необходимость автоматической сегментации, т.к. для выявления следов монтажа необходимо проверять сигналы во всех паузах.

Как было показано выше, исходя из требования обеспечения выявления монтажа по слогам, нами было экспериментально определено, что минимальная длительность сегментируемой паузы должна составлять 50 мсек, что соответствует их реальной длительности в украинской и русской речи высокого темпа. Длительность пауз между словами и отдельными фразами может достигать нескольких секунд. Таким образом, требование по минимальной длительности выделяемой паузы составляет 50 мсек.

Очевидно, что различие между уровнями, спектральным и фрактальным составам сигналов, содержащихся в паузах, могут являться основой создания системы их автоматической сегментации в фонограммах.

Далее рассмотрим, какими методами и средствами можно решить задачу выделения пауз из речи в автоматическом режиме.

Ранее нами было показано, что эта задача решается на основе применения фрактального подхода к сигналам речи. При этом автоматическое выделение сигналов пауз из речевого потока реализуется на основе метода, использующего меру Хаусдорфа [3; 4]. Но в этих работах не были показаны некоторые особенности сигналов в паузах, которые следует учесть при разработке системы автоматического выделения пауз.

К таким особенностям относятся, во-первых, наличие коротких выбросов с крутым фронтом уровней сигналов в паузе, что осложняет селекцию пауз по их истинной длительности. И, во-вторых, наличие в начале и в конце паузы малого

уровня речевого сигнала, который уже не воспринимается человеческим ухом на фоне помехи, но, тем не менее, по своим спектральным и фрактальным параметрам, никак не может быть отнесен к сигналу, отвечающему характеристикам паузы. При разработке точной автоматической системы выделения пауз эти аномалии необходимо каким-то образом блокировать. Но при этом следует помнить, что и признаки монтажа могут проявляться в виде выбросов в сигналах пауз.

Выбросы длительностью до 5 мсек при выделении пауз блокируются за счет применения метода выделения, нечувствительного к их воздействию. А выбросы большой длительности воспринимаются как участки короткого сигнала, что позволяет сохранить возможные признаки монтажа для последующего анализа. Это обеспечивается применением модифицированной формулой расчета фрактальной размерности, основанной на мере Хаусдорфа, разработанной специально для увеличения точности выделения пауз. Рассмотрим этот метод более подробно. В работе [5] фрактальная размерность по Хаусдорфу определяется, как

$$D = 2 - \lim_{p \to 0} \left[\ln N(p) \Big/ \ln(p) \right], \quad (1)$$

где:

$N(p)$ – количество ячеек при покрытии ячейками с масштабом p, которые включают хотя бы одно значение величины исследуемого сигнала,

p – масштаб разбиения.

Построим график зависимости $\ln N(p) = f(\ln p)$ при изменении масштаба для конкретной реализации временного окна паузы в фонограмме. После построения графика осуществим аппроксимацию первых 20 точек графика линейной зависимостью

$$F(p) = c \ln(p) + c_0, \quad (2)$$

где c и c_0 – коэффициенты аппроксимации, изменяющие временной масштаб [3; 5].

Величина фрактальной размерности D в рамках данного подхода равна

$$D = 2 - |c|, \quad (3)$$

где $|c|$ – абсолютная величина коэффициента c [3; 5].

Такой подход позволяет растянуть масштаб представления фрактальной меры по оси времени. Было проведено исследование фрагментов речи и пауз для нескольких сотен различных ЦФ [3]. Анализ динамики изменения фрактальной размерности фрагментов ЦФ показал, что для фрагментов речи длительностью (20 – 60) мс средняя фрактальная размерность D находится 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 указали на возможность эффективной сегментации пауз и фрагментов речи в ЦФ по параметру фрактальной размерности и были проиллюстрированы в работе [3] динамикой изменения фрактальной размерности на фрагментах речевого сигнала и сигнала шума в паузах, приведенных на рис. 1 и рис. 2. При этом длительность временного окна выбиралась равной 20 мсек, а количество точек аппроксимации пропорционально частоте дискретизации [3].

Проведенное экспериментальное исследование показало, что границей паузы и фрагмента речи является момент времени, разделяющий временное окно длительностью 20 мсек, при котором величина фрактальной размерности D переходит границу $D_q = 1{,}65$ [3].

Однако масштаб представления по оси фрактальной меры в формуле (3) несколько неудобен, что связано как с вычислением начальных и конечных участков звукового сигнала в паузах, так и с реакцией системы на выбросы сигнала внутри паузы.

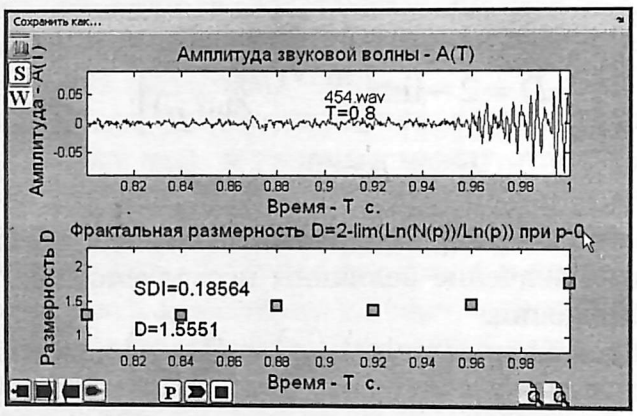

Рис. 1. Иллюстрация динамики изменения фрактальной размерности паузы

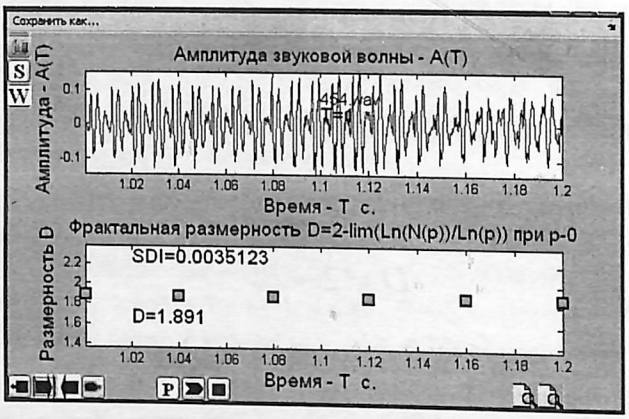

Рис. 2. Иллюстрация динамики изменения фрактальной размерности речевого фрагмента

Поэтому было предложено использовать модифицированный расчет фрактальной размерности как некой экспериментальной функции

$$f(D)=[2-|c|]f[K], \tag{4}$$

где $f(K)$ – функция переменных коэффициентов, подобранных экспериментально в процессе проведения исследований.

В результате такой модификации фрактальная размерность стала представляться графиком, приведенным на рис. 3.

Рис. 3. Фрактальная размерность паузы, рассчитанная по формуле (3)

При модифицированном расчете фрактальной размерности масштаб ее представления укрупняется и в результате речевой сигнал отождествляется со значением фрактальной меры $f_C(D) \geq 0{,}3$, а сигнал в паузе речи со значением $f_C(D) < 0{,}3$.

Использование временного окна длительностью 20 мсек позволяет выделять паузы длительностью 50 мсек и производить последующее выделение пауз с возможными признаками монтажа. Применение такой модификации обеспечило вычисление начальных и конечных участков звукового сигнала в паузах, и необходимую реакцию системы на выбросы сигнала внутри паузы, что иллюстрируется фрактальной размерностью сигнала одной из пауз, показанной на рис. 4.

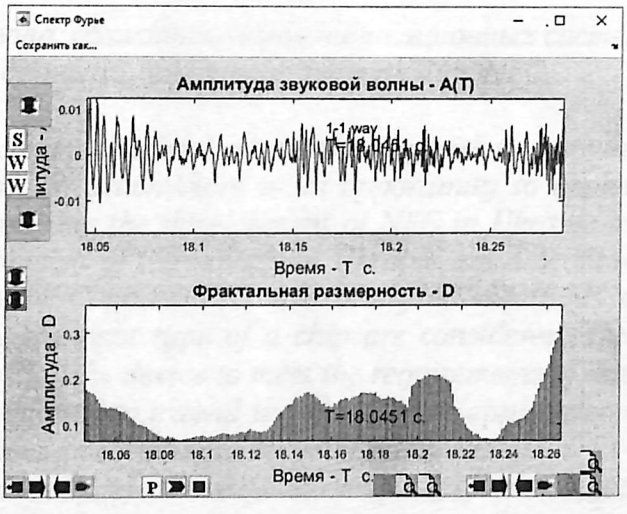

Рис. 4. Реальный сигнал паузы и его фрактальная размерность, полученная на основании модифицированной меры Хаусдорфа

Выводы

Предложен метод автоматической сегментации фонограммы по паузам, основанный на использовании модифицированной для этой задачи фрактальной размерности Хаусдорфа. Метод обеспечивает автоматическую сегментацию фонограммы на речевые сигналы и паузы при минимальной длительности выделяемых пауз 50 мсек, при этом система не реагирует на выбросы сигнала длительностью до 5 мсек внутри паузы и позволяет точно установить границы сигналов речи и пауз.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Рыбальский О.В., Жариков Ю.Ф.* Современные методы проверки аутентичности магнитных фонограмм в судебно-акустической экспертизе. К.: Нац. акад. внутр. справ України, 2003. 300 с.

2. Тлумачний словник термінів судової експертизи відеозвукозапису (українсько-російський і російсько-український) / уклад. О.Ф. Д'яченко. Х.: ХНДІСЕ, 2009. 432 с.

3. *Соловьев В.И.* Идентификация аппаратуры аудиозаписи по статистическим характеристикам аудиофайлов. Реєстрація та обробка інформації. Т. 14, 2013. № 1. С. 59–70.

4. *Журавель В.В.* Принципы, использованные при проектировании программного обеспечения проверки целостности информации в цифровых фонограммах. Сучасна спеціальна техніка. 2016. № 2(45). С. 39–43.

5. *Павлов А.Н., Анищенко В.С.* Мультифрактальный анализ сложных сигналов. Успехи физических наук. 2007. Т. 177, № 8. С. 859–876.

Отримано 17.01.2018

Рецензент Єрохін В.Ф., д.т.н., проф.