

УДК 343.98

Ю. Г. Горшков, А. М. Каиндин

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ФОНОСКОПИЧЕСКОЙ ЭКСПЕРТИЗЫ АУДИОЗАПИСЕЙ

Рассмотрен новый подход к построению инструментальных средств криминалистического исследования фонограмм. Выполнен краткий сравнительный анализ известных решений. Проведено тестирование распространенного аппаратно-программного комплекса фоноскопической экспертизы аудиозаписей “ИКАР Лаб” в составе средств ввода/вывода речевых сигналов, программного обеспечения спектрального анализа сигналов фонограмм. С учетом полученных данных рассмотрены недостатки применяемых методов спектрального анализа речевых сигналов, реализованных на основе преобразования Фурье. Предложены технические решения, обеспечивающие более высокую точность как на этапе ввода аудиозаписей в ПК, так и непосредственно при вычислении параметров речевого сигнала. Приведены примеры адаптивной многоуровневой вейвлет-обработки фонограмм, обеспечивающей получение частотно-временной структуры гласных звуков с повышенным разрешением.

E-mail: :ygorshkov@rambler.ru; alex@kaindin.ru

Ключевые слова: фоноскопическая экспертиза, преобразование Фурье, вейвлет-преобразование, речевой сигнал.

При создании средств идентификации личности по речи основное внимание разработчики уделяют измерению и наглядному отображению всех возможных акустических параметров речевых сигналов, которые могут понадобиться в ходе криминалистического исследования. Причем измерение, обработка и отображение результатов должны проводить с максимально высокой степенью точности, надежности и автоматизации.

В последние годы при исследованиях фонограмм вопросам оценки точности применяемых инструментальных средств анализа акустических сигналов аудиозаписей, а также использованию новых высокоточных методов обработки речи уделяется значительное внимание [1–5]. Надежность систем идентификации диктора по голосу также зависит от точности программных средств выделения информационных параметров речи при статистической обработке речевых сигналов.

В работе [6] показано, что окно Гаусса обеспечивает наиболее детальную картину изменения частот гармонических составляющих, на



а



б

Рис. 1. Внешний вид комплекса “ИКАР Лаб II+” (а) и устройства ввода/вывода аудиосигналов “КАМЕРТОН” (б)

которые раскладывается речевой сигнал, и является лучшим по критерию разрешающей способности по частоте и по времени. В то же время, несмотря на предлагаемые более точные решения по разложению речевого сигнала, в большинстве современных комплексов при решении задачи идентификации диктора выделение акустических признаков основывается на методе спектрального анализа с использованием быстрого преобразования Фурье (БПФ).

Аппаратно-программный комплекс фonoскопической экспертизы аудиозаписей “ИКАР Лаб II+”. В настоящее время у экспертов-фonoскопистов широкое применение находит аппаратно-программный комплекс криминалистического исследования аудиозаписей “ИКАР Лаб II+” [7] компании “Центр речевых технологий” (Санкт-Петербург), в состав которого входит устройство для измерения характеристик и формирования электрических сигналов в звуковом диапазоне частот STC-H246 “КАМЕРТОН” [8]. На рис. 1 представлен внешний вид комплекса “ИКАР Лаб II+” и устройства “КАМЕРТОН”. Основные технические характеристики комплекса: число аудио-каналов записи/воспроизведения — 2 моно или 1 стерео; динамический диапазон аудио — не менее 105 дБ.

Комплекс “ИКАР Лаб II+” предназначен для записи, анализа, шумочистки и установления подлинности фонограмм. Принятие решения по идентификации личности по голосу в перечисленных сред-

ствах основывается на оценке спектральных характеристик гласных звуков, полученных на основе преобразования Фурье. Вычисление спектрально-формантных характеристик и динамических спектрограмм частотно-временного представления речевого сигнала осуществляется с использованием БПФ.

В целях получения реальных характеристик тракта аудиоввода комплекса “ИКАР Лаб II+” выполнено тестирование устройства “КАМЕРТОН”. Получены следующие данные: неравномерность амплитудно-частотной характеристики (АЧХ) в частотном диапазоне от 40 Гц до 15 кГц составляет +0,01 дБ, -0,01 дБ; уровень шума 82,5 дБ; динамический диапазон 82,3 дБ; нелинейные искажения 0,0034 %; интермодуляционные искажения 0,030 %.

При проведении криминалистической экспертизы множество акустических признаков голоса и речи исследуемого диктора может быть представлено в виде совокупности фонетических (акустических) классов: гласные звуки (ударные, безударные) и согласные (назальные, аффрикаты, щелевые и др.). Эти акустические классы отражают артикуляторные особенности речи диктора, а также конфигурацию его речевого тракта.

Известные методы анализа звуков речи основываются на спектральной модели стационарного сигнала [9]. Основным недостатком этой модели является отсутствие характеристик для шумовых составляющих в произносимых согласных звуках, и это при том, что в большинстве языков основная речевая информация передается согласными. Традиционно разрабатываемые алгоритмы идентификации личности по голосу и распознавания речи основываются, как правило, на определении характеристик гласных звуков.

Разработанная специалистами Центра речевых технологий методика криминалистической идентификации дикторов по речи на русском и других языках [10], является развитием известной методики идентификации дикторов “Диалект” [11] и также включает в себя на этапе инструментального анализа сравнение статистик основного тона голоса и формант, формантное выравнивание, экспертное сравнение формант ударных гласных.

На рис. 2 приведен один из типичных примеров формантного сравнения речи двух дикторов, приводимых в качестве иллюстрации к заключению эксперта [10].

Анализ речевого сигнала с использованием преобразования Фурье. Преобразование Фурье является классическим методом частотного анализа стационарных сигналов непрерывного времени (СТФТ). При этом сигнал раскладывается в базис синусов и косинусов различных частот. Коэффициенты преобразования находятся

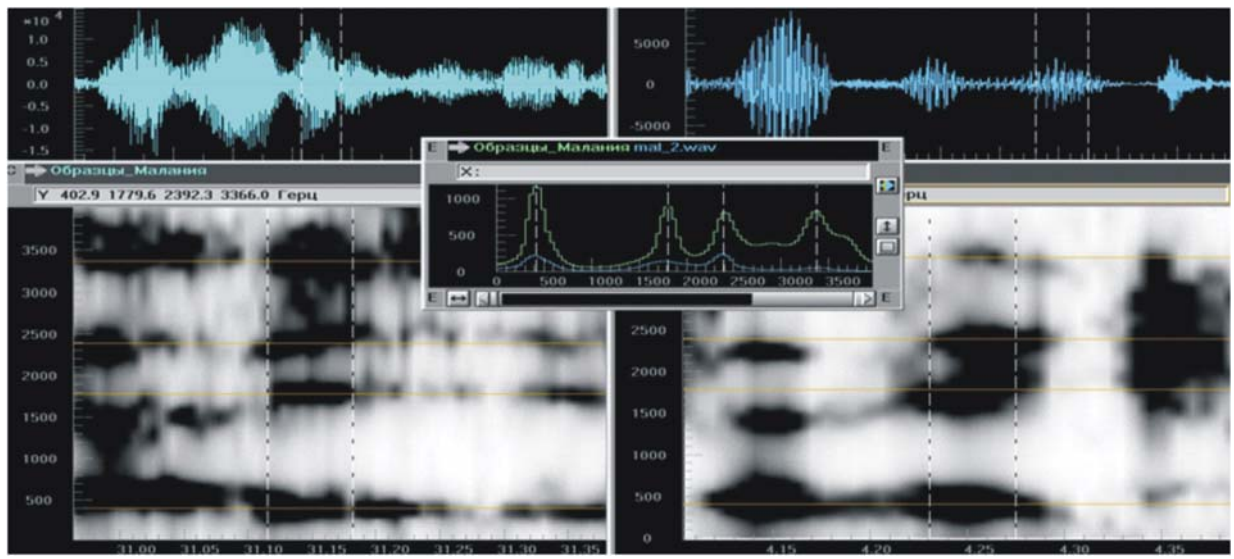


Рис. 2. Пример временных, амплитудных и спектрально-формантных характеристик произнесения гласного [i] для речи сравниваемых дикторов. В средней части рисунка изображены усредненные спектры участков сопоставляемых звуков

путем вычисления скалярного произведения сигнала с комплексными экспонентами:

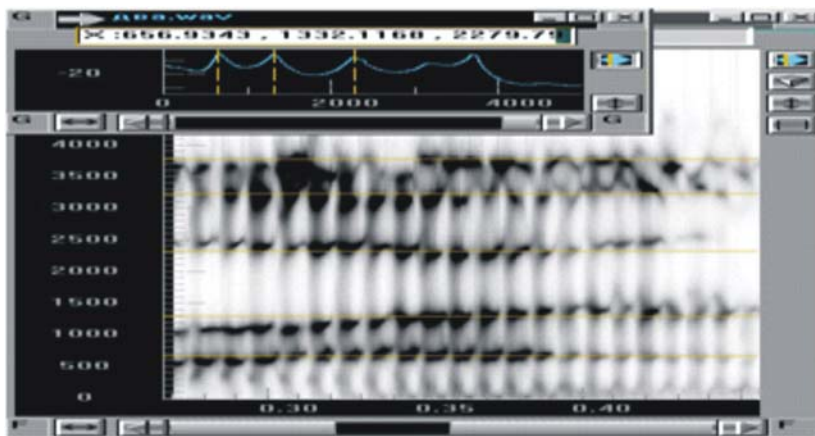
$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x)e^{-i\omega x} dx. \quad (1)$$

Результат преобразования Фурье — амплитудно-частотный спектр, по которому можно определить присутствие некоторой частоты в исследуемом сигнале.

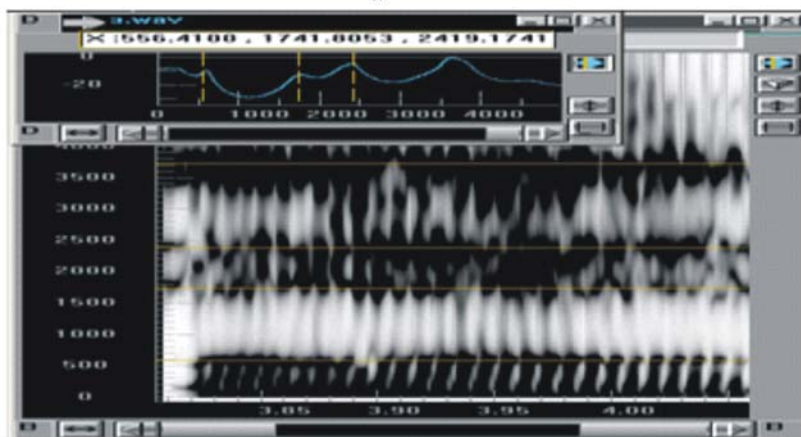
С практической точки зрения СТФТ имеет ряд недостатков. Во-первых, для получения преобразования на одной частоте требуется вся временная информация. Это означает, что должно быть известно будущее поведение сигнала. На практике не все сигналы стационарны. Пик сигнала во временной области распространится по всей частотной области его преобразования Фурье. Для преодоления этих недостатков СТФТ вводится кратковременное, или оконное, преобразование Фурье (STFT):

$$F(\omega, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x)e^{-i\omega x} w(x - b) dx, \quad (2)$$

в котором применяется операция умножения сигнала на окно перед применением преобразования Фурье. Окном $w(x - b)$ является локальная функция, которая сдвигается вдоль временной оси для вычисления преобразования в нескольких позициях b . Преобразование становится зависимым от времени, и в результате получается частотно-временное описание сигнала. Данный подход позволяет определить факт присутствия в сигнале любой частоты и интервал ее присутствия. Это



a



б

Рис. 3. Фурье-сонограмма звуков “а” (а) и “э” (б)

значительно расширяет возможности метода по сравнению с классическим преобразованием Фурье. В качестве окна часто выбирается функция Гаусса.

На рис. 3 представлены фурье-сонограммы или изображения “видимый звук” гласных “а” и “э” с использованием комплекса “ИКАР Лаб II+”. Ввод речевого сигнала в ПК осуществлялся с помощью устройства “КАМЕРТОН”, при построении сонограмм применялась программа SIS (вычисление спектра на основе БПФ).

Недостаток STFT состоит в том, что при его вычислении используется фиксированное окно, которое не может быть адаптировано к локальным свойствам сигнала. Согласно принципу неопределенности Гейзенберга в данном случае нельзя утверждать факт наличия частоты ω_0 в сигнале в момент времени t_0 , — можно лишь определить, что спектр частот (ω_1, ω_2) присутствует в интервале (t_1, t_2) . Причем разрешение в координатах время–частота остается постоянным.

Речевой сигнал является примером нестационарного процесса, в котором информативным является сам факт изменения его частотно-временных характеристик. Для выполнения анализа таких процессов требуются базисные функции, имеющие способность выявлять в анализируемом сигнале как частотные, так и его временные характеристики. Другими словами, сами функции должны обладать свойствами частотно-временной локализации.

Вейвлет-преобразование имеет такие свойства, а также ряд существенных преимуществ при выполнении высокоточного анализа сложных нестационарных сигналов, к которым следует отнести исследуемые аудиозаписи фонограмм.

Вейвлет-анализ речевого сигнала. Вейвлет-преобразование сигналов является обобщением спектрального анализа. Применяемые для этой цели базисы были названы вейвлетами – функциями двух аргументов: масштаба и сдвига. В отличие от преобразования Фурье, вейвлет-преобразование обеспечивает двумерное представление исследуемого сигнала в частотной области в плоскости частота–положение. Аналогом частоты при этом является масштаб аргумента базисной функции (чаще всего времени), а положение характеризуется ее сдвигом. Это позволяет разделять крупные и мелкие особенности сигналов, одновременно локализуя их на временной шкале. Иными словами, вейвлет-анализ можно охарактеризовать как спектральный анализ локальных возмущений [12]. Практические вопросы применения теории “всплесков” (вейвлет-преобразования) для получения сонограмм с частотно-временным разрешением, обеспечивающим выделение параметров не только гласных, но и согласных звуков приведены в работе [13].

Аппаратно-программные средства исследования фонограмм с использованием технологии вейвлет-анализа. В ходе поисковых исследований на кафедре “Информационная безопасность” МГТУ им. Н.Э. Баумана для отладки программных средств высокоточной обработки речевых сигналов создан исследовательский комплекс “Фон” частотно-временного анализа с использованием вейвлет-технологии. Возможности комплекса позволили довести экспериментальное программное обеспечение вейвлет-анализа речи [14] и средства ввода/вывода сигналов до уровня специализированных аппаратно-программных средств высокоточной регистрации аудиосигналов и многоуровневого вейвлет-анализа фонограмм [15, 16].

Достоверность выполненной экспертизы фонограммы во многом зависит от точности средств регистрации речевых сигналов, включающей АЧХ тракта записи, отношение сигнал/шум, динамический диапазон, нелинейные и интермодуляционные искажения. Следует отметить, что при сравнении результатов тестирования устройства “КАМЕРТОН” и данных исследования модуля записи NI PXI-4461 (National Instruments, США) основные технические характеристики последнего значительно превышают параметры отечественного изделия.

На рис. 4 показан модуль NI PXI-4461. Результаты тестирования модуля записи NI PXI-4461; 24 бита; 204,8 кС/с; динамический диапазон 118 дБ; неравномерность АЧХ в диапазоне от 40 Гц до 15 кГц со-



Рис. 4. Внешний вид модуля NI PXI-4461

ставляет +0,00 дБ, –0,03 дБ; уровень шума 7,7 дБ; динамический диапазон 97,8 дБ; нелинейные искажения 0,0008 %; интермодуляционные искажения 0,0071 %. На рис. 5 приведена структура специализированного аппаратно-программного комплекса “Фон-НИ”.

На рис. 6, а и б представлены вейвлет-сонограммы гласных “а” и “э”, полученные с использованием специализированного аппаратно-программного комплекса “Фон-НИ” [16], в состав которого входит модуль NI PXI-4461 и специальное программное обеспечение WaveView-4 [13].

Из анализа изображений сонограмм рис. 6 и рис. 3 следует, что вейвлет-сонограммы, полученные с использованием программы WaveView-4 и модуля NI PXI-4461 обеспечивают более высокое частотно-временное разрешение тональных участков речевого сигнала (гласных звуков).

Заключение. Комплексы криминалистического исследования аудиозаписей являются основным инструментом экспертов-фоноскопистов при решении сложных задач идентификации личности по речевому сигналу. Вычисление наиболее значимых параметров речевого сигнала – спектрально-формантных характеристик гласных звуков – для последующего сравнения и принятия решения экспертом о принадлежности исследуемых записей данному диктору программными средствами комплексов осуществляется с использованием классического преобразования Фурье.

В настоящей работе предложен новый подход в создании аппаратно-программных средств, обеспечивающих более точную, по сравнению с известными решениями обработку нестационарных сигналов речевого диапазона. Проведенное тестирование аппаратно-программных средств анализа фонограмм показало, что наилучшую точность на

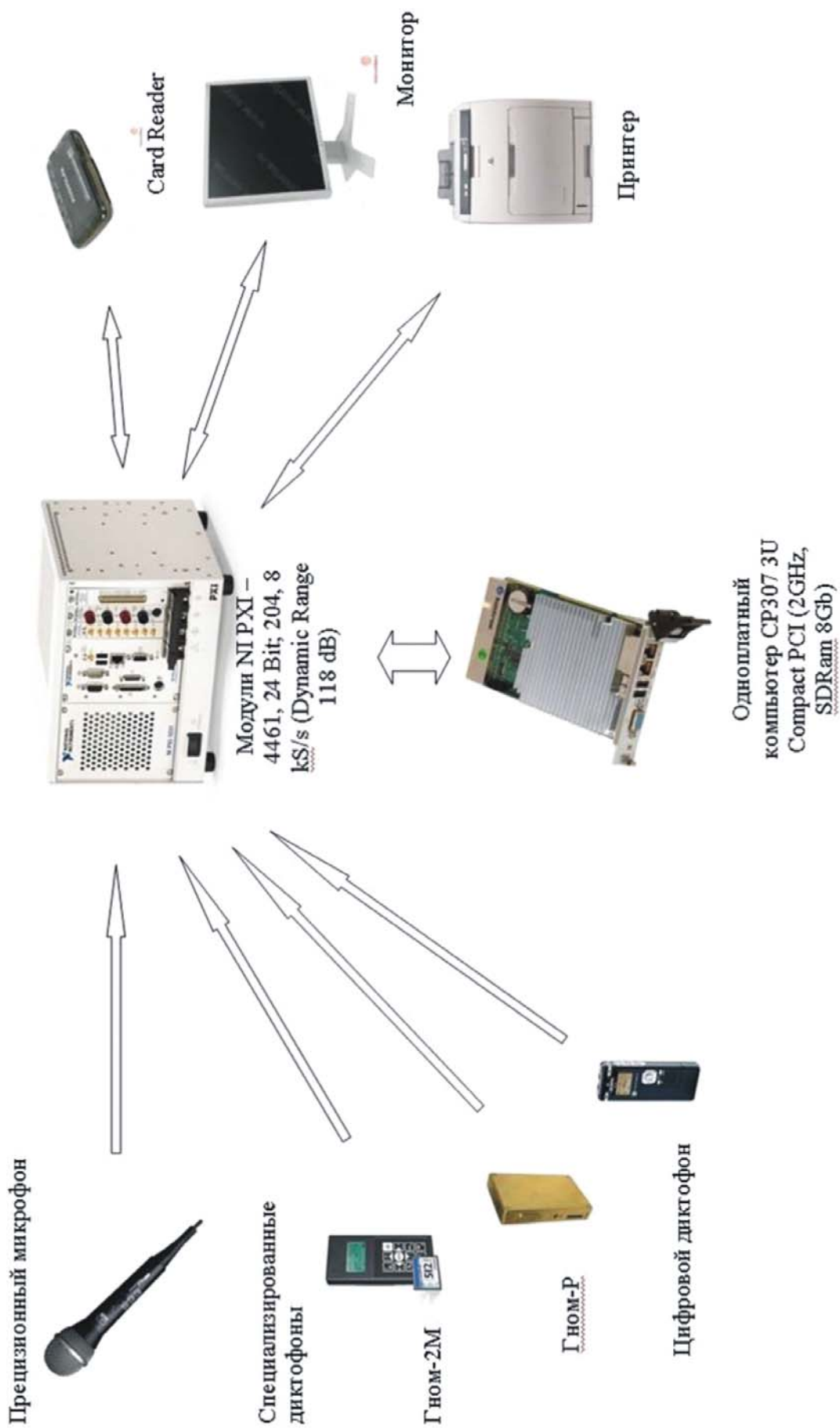
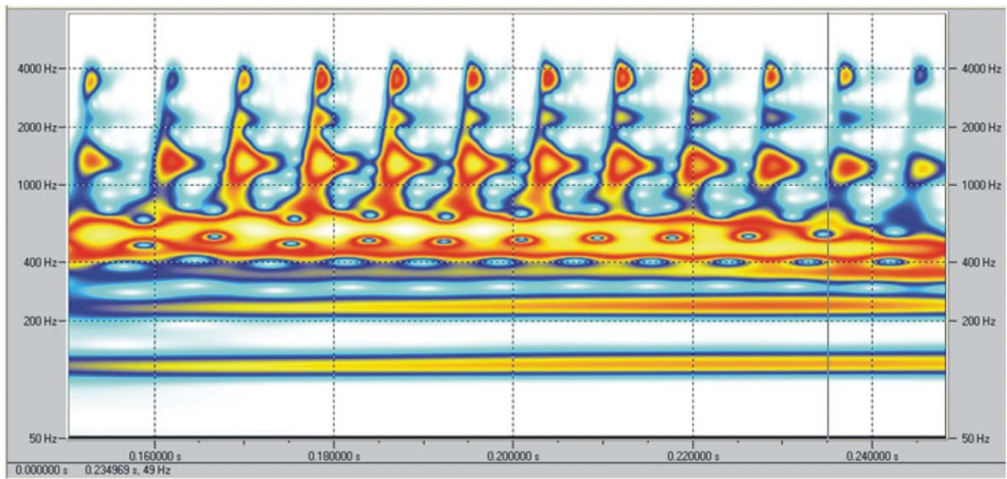
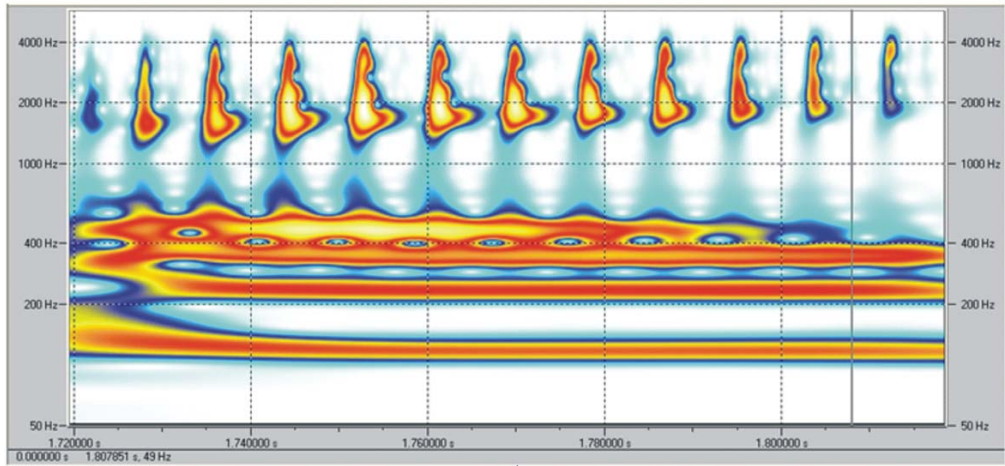


Рис. 5. Структура специализированного аппаратно-программного комплекса “Фон-PI”



а



б

Рис. 6. Вейвлет-сонограмма звука “а” (а) и “э” (б)

этапе ввода сигналов обеспечивают модули записи NI PXI-4461 компании National Instruments. Данные модули по своим техническим характеристикам значительно превосходят возможности устройства “КАМЕРТОН” — ввода сигналов в ПК аппаратной части наиболее распространенного в отечественных коммерческих экспертных организациях комплекса криминалистического исследования фонограмм “ИКАР Лаб II+”.

Обработка аудиозаписей 440 дикторов (мужчины и женщины в возрасте 22...24 года) показала, что на гласных русского языка при многоуровневом вейвлет-анализе проявляется более тонкая частотно-временная структура. За счет более точной обработки высокочастотная часть (область общепринятых 3 и 4 формант) тональных участков речевого сигнала представляет собой последовательности сложных нестационарных пульсаций.

Разработанный аппаратно-программный комплекс криминалистического исследования фонограмм “Фон-НИ” в составе модулей NI PXI-4461 и программных средств обработки речевых сигналов на основе многоуровневого вейвлет-преобразования WaveView-4 подготовлен для проведения сертификационных испытаний на соответствие международным требованиям.

Применение комплексов фоноскопической экспертизы “Фон-НГ” позволит повысить точность обработки аудиозаписей и, следовательно, достоверность полученных результатов экспертов при проведении криминалистического исследования аудиозаписей.

СПИСОК ЛИТЕРАТУРЫ

1. Ж е н и л о В. Р. Компьютерная фоноскопия. – М.: Академия МВД России, 1995. – 208 с.
2. Г а л я ш и н а Е. И. Судебная фоноскопическая экспертиза. – М.: Триада, 2001. – 176 с.
3. Г о р ш к о в Ю. Г. Аппаратно-программные средства анализа, шумоочистки и засекречивания речевого сигнала коммерческого применения (3-е поколение: вейвлет-технологии) // Тез. докл. 1-й Моск. междунар. конф. “Интегрированные системы безопасности: новейшие технологии”, 26–27 апреля 2004 г. – Москва. – С. 2.
4. Г о р ш к о в Ю. Г. Многоуровневый вейвлет-анализ акустических сигналов при решении задач фоноскопической экспертизы // Информатизация и информационная безопасность правоохранительных органов: Материалы 20-й Междунар. науч. конф. – М., 2011. – С. 379–387.
5. Г о р ш к о в Ю. Г. Инструментальное исследование фонограмм с использованием программных средств многоуровневого вейвлет-анализа сигналов // Труды Междунар. конгресса по информатике: информационные системы и технологии (International Congress on Computer Science: Information Systems and Technologies, CSIST’2011) 31 октября – 3 ноября 2011 г. – Минск. – С. 131–135.
6. Ж е н и л о В. Р. Информация, звук и преобразование Фурье–Гаусса // Информатизация и информационная безопасность правоохранительных органов: Материалы 15-й Междунар. науч. конф. – М., 2006. – С. 332–338.
7. А п п а р а т н о-программный комплекс криминалистического исследования фонограмм речи “ИКАР Лаб II+”. Формуляр. – СПб.: Центр речевых технологий, 2006.
8. У с т р о й с т в о для измерения характеристик и формирования электрических сигналов в звуковом диапазоне частот STC-H246 “КАМЕРТОН”: Руководство по эксплуатации. – СПб.: Центр речевых технологий. – 2006.
9. Ф а н т Г. Акустическая теория речеобразования. (G. Fant. Acoustic theory of speech production, 1960) / Пер. с англ. Л.А. Варшавского и В.И. Медведева: Под ред. В.С. Григорьева. – М.: Наука, 1964. – 284 с.
10. К о в а л ь С. Л. Методика идентификации дикторов по голосу и речи на основе комплексного анализа фонограмм. – СПб.: Центр речевых технологий, 2006. – 18 с.
11. Т и м о ф е е в Е. Н., Г о л о щ а п о в а Т. И., Д о к у ч а е в И. В. Применение автоматизированной системы “Диалект” на базе компьютерной речевой лаборатории CSL (США) при решении задач идентификации дикторов: Учеб. пособие. – М.: ЭКЦ МВД России, 2000. – 120 с.
12. Д ь я к о н о в В. И. Вейвлеты: от теории к практике. – М.: Солон-Р, 2002.
13. Г о р ш к о в Ю. Г. Исследовательский комплекс частотно-временного анализа речевого сигнала с использованием вейвлет-технологии // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. – 2011. – № 4. – С. 78–87.
14. Г о р ш к о в Ю. Г., К у з н А. Ю. Применение Wavelet-преобразования при решении задач анализа речевого сигнала // Сб. трудов X Всеросс. науч. конф. Проблемы информационной безопасности в системе высшей школы. – М., 2003. – С. 24.

15. Горшков Ю. Г., Пестряков А. А. Специализированные средства анализа речевых сигналов с использованием вейвлет-преобразования // Проблемы информационной безопасности в системе высшей школы. XV Всеросс. науч.-практич. конф. – М., 2008. – С. 36.
16. Горшков Ю. Г. Специализированные технические средства регистрации речевого сигнала при решении задач фоноскопической экспертизы // XIX Междунар. науч.-практич. конф. Образовательные, научные и инженерные приложения в среде LabVIEW и технологии National Instruments–2010, Секция – Научно-исследовательские и испытательные стенды, 3–4 декабря 2010 г.: Тез. докл. – М. – С. 353–355.

Статья поступила в редакцию 27.02.2012

Юрий Георгиевич Горшков родился в 1945 г., окончил в 1969 г. Новосибирский электротехнический институт связи. Канд. техн. наук, доцент кафедры “Информационная безопасность” МГТУ им. Н.Э. Баумана. Автор более 50 научных работ в области информационной безопасности и разработки защищенных систем связи.

Yu.G. Gorshkov (b. 1945) graduated from the Novosibirsk Electrotechnical Communication Institute in 1969. Ph. D. (Eng.), assoc. professor, head of “Information Security” department of the Bauman Moscow State Technical University. Author of more than 50 publications in the field of data security and development of protected communication systems.

Александр Михайлович Каиндин родился в 1989 г. Студент кафедры “Информационная безопасность” МГТУ им. Н.Э. Баумана. Автор ряда научных работ в области информационной безопасности.

A.M. Kaindin (b. 1989) — student of “Information Security” department of the Bauman Moscow State Technical University. Author of more than 50 publications in the field of data security.