

ИНСТРУМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ФОНОГРАММ С ИСПОЛЬЗОВАНИЕМ ПРОГРАММНЫХ СРЕДСТВ МНОГОУРОВНЕВОГО ВЕЙВЛЕТ-АНАЛИЗА СИГНАЛОВ

Ю. Г. Горшков

*Московский государственный технический университет
имени Н. Э. Баумана*

Москва, Россия

E-mail: YGorshkov@rambler.ru

В статье делается сравнительный анализ аппаратно-программных средств криминалистического исследования фонограмм. Рассмотрены преимущества вейвлет-преобразования речевых сигналов по сравнению с преобразованием Фурье и возможности специального программного обеспечения WaveView-4 высокоточного многоуровневого вейвлет-анализа. Представлены частотно-временные изображения-вейвлет-сонограммы гласных и согласных звуков, а также приводится пример установления факта «монтажа» фонограммы.

Ключевые слова: исследование фонограмм, вейвлет-анализ, речевой сигнал.

При проведении криминалистического исследования фонограмм вопросам оценки точности используемых инструментальных средств анализа акустических сигналов аудиозаписей, а также использованию новых высокоточных методов анализа уделяется значительное внимание [1, 2, 3]. Надежность систем идентификации диктора по голосу также зависит от точности программных средств выделения информационных параметров речи при статистической обработке акустических сигналов.

В работе [4] показано, что наиболее детальную картину изменения частот гармонических составляющих, на которые раскладывается речевой сигнал, и лучшим по критерию разрешающей способности по частоте и по времени обеспечивает окно Гаусса. 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-ПРОГРАММНЫЕ КОМПЛЕКСЫ ИССЛЕДОВАНИЯ ФОНОГРАММ

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.

Большинство современных методов анализа звуков речи основываются на спектральной модели для стационарного сигнала [5]. Основным недостатком этой модели является отсутствие характеристик для основных 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 (Санкт-Петербург) методика криминалистической идентификации дикторов по речи на русском и других языках [6] является развитием известной методики идентификации дикторов «Диалект» и также включает в себя на этапе инструментального анализа сравнение статистик основного тона голоса и формант, формантное «выравнивание», экспертное сравнение формант ударных гласных.

На рисунке 1 представлен один из типичных примеров формантного сравнения речи двух дикторов, приводимых в качестве иллюстрации к заключению эксперта [6].

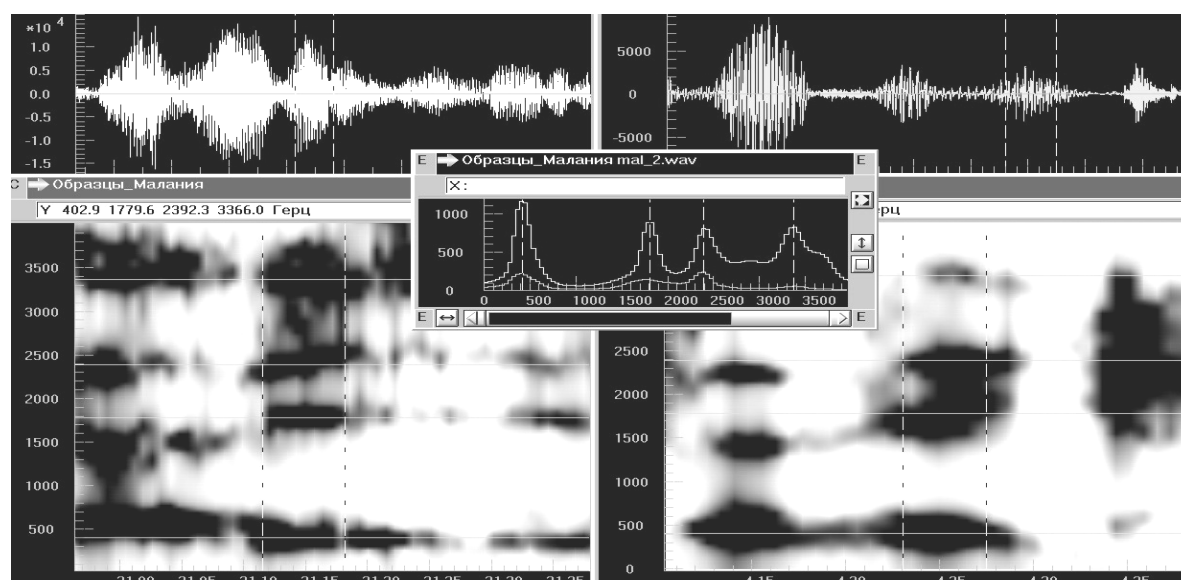


Рис. 1. Пример временных, амплитудных и спектрально-формантных характеристик произнесения гласного [i] для речи сравниваемых дикторов. В средней части рисунка изображены усредненные спектры участков сопоставляемых звуков

Помимо системы «Диалект», в последние годы широкое применение находят комплексы исследования фонограмм «ИкарЛаб», «Justiphone», «ОТExpert», разработанные российскими компаниями. Принятие решения по идентификации личности по голосу в перечисленных средствах основывается на оценке спектральных характеристик гласных звуков, полученных на основе преобразования Фурье.

На этапе предварительного анализа фонограмм используются звуковые редакторы: Adobe Audition, AWave, Cool Edit, Sound Forge, Speech Analyzer, Steinberg WaveLab, Wave Flow, WaveLab, в которых частотно-временные представления речевого сигнала или динамические спектрограммы также строятся с использованием алгоритма БПФ.

ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

В последние годы вейвлеты находят широкое применение при фильтрации, предварительной обработке и синтезе различных сигналов, решении задач сжатия и обработки изображений. Особенно важна принципиальная возможность вейвлетов представлять нестационарные сигналы [7].

Все большее число специалистов по цифровой обработке сигналов убеждаются в том, что преобразования Фурье в классическом виде не обеспечивают необходимую точность представления нестационарных сигналов, в частности, к которым относятся речевые сигналы. Значительный интерес представляют утверждения о том, что вейвлет-спектрограммы намного более информативны, чем обычные Фурье-спектрограммы, и (в отличие от последних) позволяют легко выявлять тончайшие локальные особенности акустических сигналов.

Вейвлеты (wavelets) – это обобщенное название особых функций, имеющих вид коротких волновых пакетов с нулевым интегральным значением и с той или иной, подчас очень сложной, формой, локализованных по оси независимой переменной (t или x) и способных к сдвигу по ней и масштабированию (сжатию/растяжению).

Вейвлеты создаются с помощью специальных базовых функций – прототипов, задающих их вид и свойства и удовлетворяющих целому ряду условий. Набор вейвлетов в их временном или частотном представлении может приближать сложный сигнал или изображение, причем идеально точно или с некоторой погрешностью.

Вейвлеты, как средство многомасштабного анализа, позволяют выделять одновременно как основные характеристики сигнала, так и короткоживущие высокочастотные явления в речевом сигнале. Это свойство является существенным преимуществом в задачах обработки речевого сигнала по сравнению с оконным преобразованием Фурье, где, варьируя ширину окна, приходится выбирать масштаб явлений, которые необходимо выделить в сигнале.

В ходе исследований, проводимых на кафедре «Информационная безопасность» МГТУ имени Н. Э. Баумана, создавалось специальное программное обеспечение (СПО) WaveView высокочастотного вейвлет-анализа с последующим построением изображений «видимый звук» или сонограмм. Первая версия СПО WaveView-1 разработана в 2002 г. [8]. WaveView-4 является последней версией ПО семейства вейвлет-анализа сигналов, в котором реализован алгоритм многоуровневой обработки [9].

На рисунке 2 для сравнения представлены вейвлет- и Фурье-сонограммы слов: «раз», «два», «три», «четыре», «пять», «шесть».

Речевые сигналы следует относить к сложным нестационарным сигналам, в связи с этим, как видно из сравнения вейвлет- и Фурье-сонограмм, применение вейвлет-преобразования позволяет получать более точное его частотно-временное представление. Преимущество вейвлет-преобразования перед преобразованием Фурье заключается в том, что оно позволяет хорошо локализовать как низкочастотные, так и высокочастотные составляющие сигнала. Отмеченные особенности нашли применение при оценке эмоционального состояния человека по его речи [10], а также при построении системы засекречивания телефонных переговоров нового поколения [11].

На рисунке 3 представлена многоуровневая вейвлет-сонограмма слова *шесть*.

На рисунке 4 представлен пример многоуровневого вейвлет-анализа сигнала фонограммы с целью определения ее подлинности. Высокое частотно-временное разрешение представленного ниже сигнала участка фонограммы позволило установить факт «монтажа».

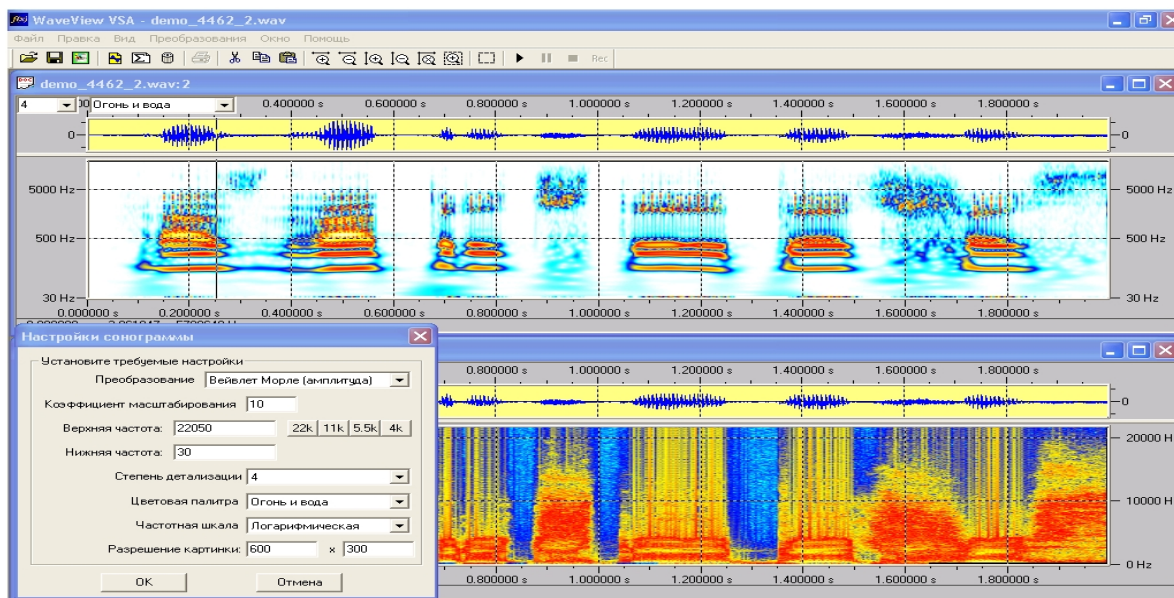


Рис. 2. Сравнение сонограмм. Верхнее изображение: многоуровневая вейвлет-сонограмма; нижнее: Фурье-сонограмма; слева внизу – меню СПО WaveView-4

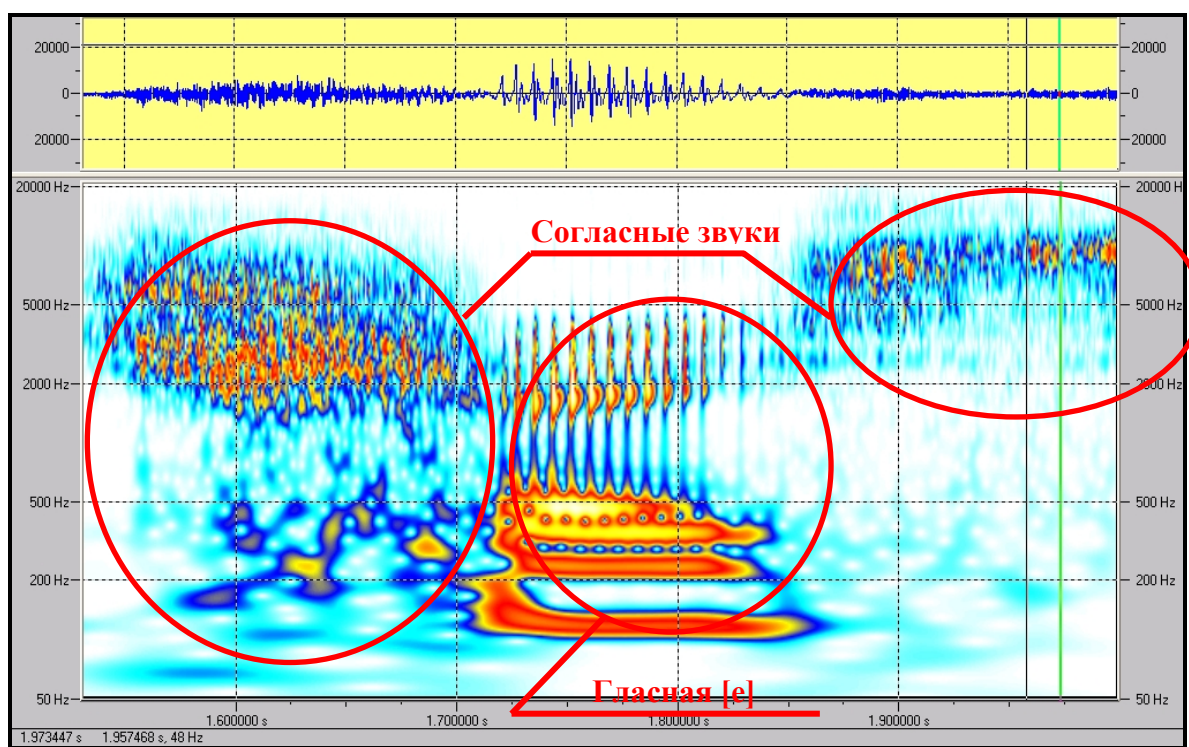


Рис.3. Многоуровневая вейвлет-сонограмма слова *шесть*

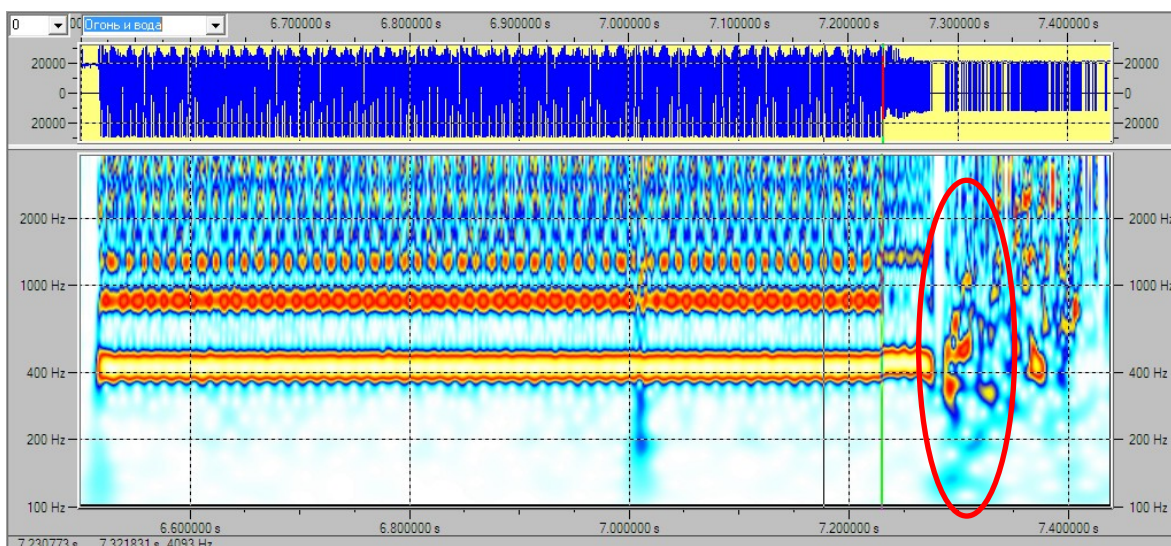


Рис.4. Многоуровневая вейвлет-сонограмма начала сеанса телефонной связи. Длительность тональной посылки «вызова» составляет 0.768 с (6.51 с – 7.278 с), что на 0.232 с меньше ее стандартной длины 1 с. В конце тональной посылки выявлен участок «монтажного перехода»; сигнал на интервале 7.22 с – 7.278 с не является его истинным продолжением, а «состыкован» с тональным сигналом, взятым из другого сеанса связи

ЛИТЕРАТУРА

1. Женило, В. Р. Компьютерная фоноскопия / В. Р. Женило. М.: Академия МВД России, 1995. 208 с.
2. Галяшина, Е. И. Судебная фоноскопическая экспертиза / Е. И. Галяшина. М. : Триада, 2001. 176 с.
3. Горшков, Ю. Г. Многоуровневый вейвлет-анализ акустических сигналов при решении задач фоноскопической экспертизы / Ю. Г. Горшков // Информатизация и информационная безопасность правоохранительных органов: материалы 20 Междунар. науч. конф. М., 2011. С. 379–387.
4. Женило, В. Р. Разложение звукового сигнала на сонели / В. Р. Женило, М. В. Женило, Д. Н. Калюжный // «Информатизация и информационная безопасность правоохранительных органов: материалы 16 Междунар. науч. конф. М., 2007. С. 91–95.
5. Фант, Г. Акустическая теория речеобразования / Г. Фант. М. : Наука, 1964. 284 с.
6. Коваль, С. Л. Методика идентификации дикторов по голосу и речи на основе комплексного анализа фонограмм / С. Л. Коваль [и др.]. СПб. : Центр Речевых Технологий, 2006. 18 с.
7. Дьяконов, В. И. Вейвлеты: от теории к практике / В. И. Дьяконов. М. : Солон-Р, 2002. 440 с.
8. Горшков, Ю. Г. Применение Wavelet-преобразования при решении задач анализа речевого сигнала / Ю. Г. Горшков, А. Ю. Кузин // Проблемы информационной безопасности в системе высшей школы: сб. тр. X Всерос. науч. конф. М., 2003. С. 24.
9. Горшков, Ю. Г. Специализированные средства анализа речевых сигналов с использованием вейвлет-преобразования / Ю. Г. Горшков, А. А. Пестряков // Проблемы информационной безопасности в системе высшей школы: сб. тр. XV Всерос. науч.-практ. конф. М., 2008. С. 36.
10. Горшков, Ю. Г. Новые решения речевых технологий безопасности / Ю. Г. Горшков // Специальная техника. 2006. № 4. С. 41–47.
11. Горшков, Ю. Г. Анализ и засекречивание речевого сигнала: учеб. пособие / Ю. Г. Горшков. М. : МГТУ имени Н. Э. Баумана, 2007. 34 с.