

Features Speech Signature Image Recognition on Mobile Devices

Key words: speech signature, audiomarker, spectrogram image.

The algorithms for dynamic spectrograms images recognition, processing and sound–speech signature (SS) were developed. The software for mobile phones, that can recognize speech signatures was prepared. The investigation of the SS recognition speed on its boundary types was conducted. Recommendations on the boundary types choice in the optimal ratio of recognition speed and required space were given.

A.M. Алюшин, Н.С. Дворянкин

ОСОБЕННОСТИ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ РЕЧЕВОЙ ПОДПИСИ НА МОБИЛЬНЫХ УСТРОЙСТВАХ

Введение

В настоящее время существует ряд важных задач, связанных с повышением защищенности информации, в частности повышением защищенности бумажных документов (договоров, соглашений, ценных бумаг и т.п.), объем оборота которых в современном мире, несмотря ни на что, не уменьшается [1 – 3].

Те способы, которые используются в настоящее время для проверки подлинности документов, не идеальны, при желании легко имитируются. Ни для кого не секрет, что можно подделать подпись, стоящую под документом, водяные знаки, несанкционированно использовать печать и т.д. Значительно повысить защищенность документов от подделки позволяет как применение технологий, основанных на личных биометрических данных человека, поскольку они сугубо индивидуальны, так и использование информационной связи между маркером и защищаемым текстом, при которой изменение текста приводит к изменению маркера и наоборот. В этой области одной из новых универсальных и экономичных технологий является использование речевых сообщений в качестве своеобразных аудиомаркеров, смысловое содержание которых, понимаемое после их озвучивания, связано с защищаемым документом, а уникальные голосовые признаки подтверждают авторство исполнителя документа [1].

Описание технологии и условий реализации на мобильных устройствах

Итак, сам способ защиты заключается в следующем. В конец или любое другое место текстового документа добавляется рисунок речевой подписи. Как правило, это изображение узкополосной спектрограммы речевого сигнала, обрамленное контурами границ, которое можно отсканировать в компьютер или снять на камеру мобильного телефона для распознавания, синтеза и озвучивания. Как уже отмечалось, в РП может храниться некоторая важная, значимая информация, связанная с исходным текстом таким образом, что изменение основной части текста должно повлечь за собой изменение речевой подписи, что сделать без автора (исполнителя) этого документа невозможно.

Несоответствие содержания озвученной фонограммы РП значимым информационным разделам документа является признаком его подмены, фальсификации, а совпадение – признаком его подлинности.

На рис. 1 приведен пример речевой подписи. Границы рисунка – параллельные черные линии сверху и снизу – служат для идентификации рисунка на смартфоне. Сам графический образ представляет собой частотно-временную спектральную развертку РС, где по оси абсцисс идет временная, а по оси ординат частотная шкала, а в уровнях

серого цвета отмечена мощность сигнала на данной частоте в данный момент времени (черный цвет – максимум энергии, белый – минимум).

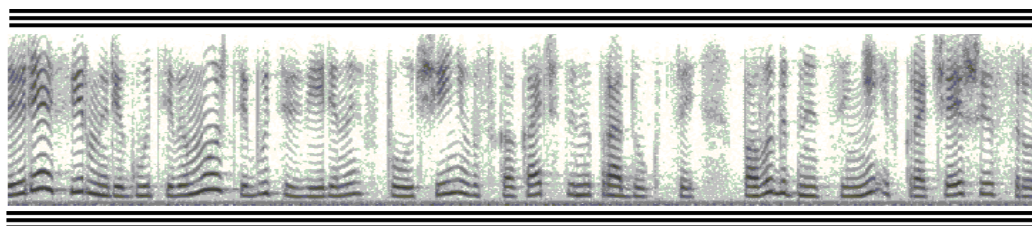


Рис. 1. Пример речевой подписи длительностью 12 с

Под важной информацией, помещаемой в РП, подразумеваются некоторые данные, которые подчеркивают основные моменты содержания текстового документа, как, например, договор, сумму договора, обязанности сторон, сроки, номера счетов, телефоны, адреса и т.д.

Общая схема использования технологии речевой подписи представлена на рис. 2.

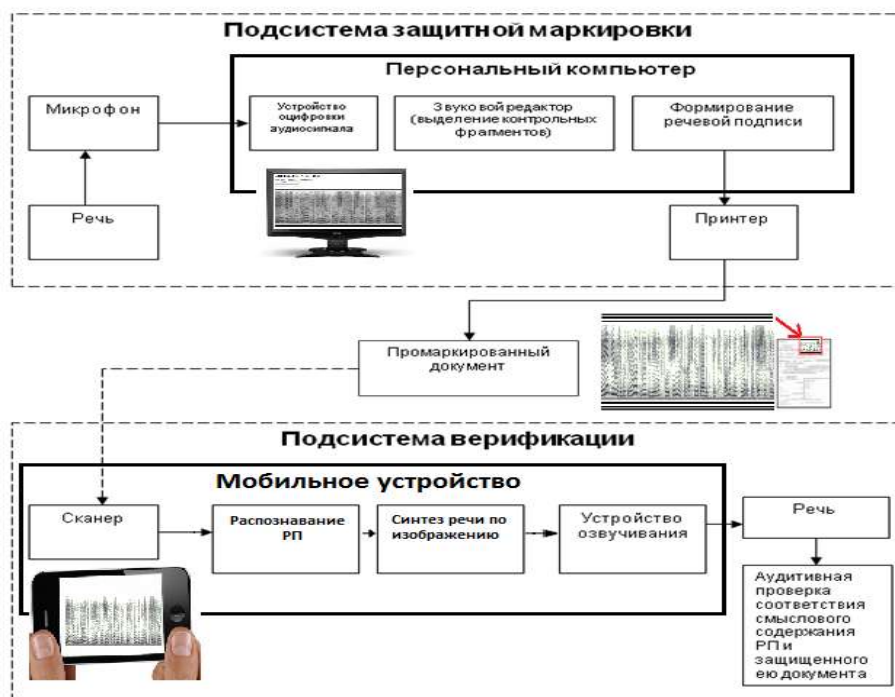


Рис. 2. Схема использования технологии речевой подписи

Предполагается, что процесс формирования РП и нанесения ее на бумажные или иные документы и изделия реализуется известным ПО для широко распространенных средств офисной техники (компьютеры со звуковым вводом/выводом, принтеры), его основные процедуры описаны в работе [1]. Та же техника, плюс сканер использовались ранее для считывания, распознавания и озвучивания РП, прикрепленной к носителю информации (бумажный документ, изделие, деталь и др.) в виде либо непосредственного изображения, либо «стикера» с ним.

К сожалению, в силу отсутствия до настоящего времени соответствующего программного обеспечения для мобильной работы с речевой подписью, эта технология пока не получила широкомасштабного распространения в различных областях своего

возможного применения, в том числе защитного маркирования документов и изделий для борьбы с контрафактом.

В связи с этим, задача создания программного комплекса аудиомаркирования носителей информации и изделий для повышения уровня защищенности содержащихся в них сведений от модификации и подмены и его реализация на мобильных устройствах-сотовой связи весьма актуальна.

В данной работе представляется разработанный авторами программный комплекс распознавания, обработки и озвучивания речевой подписи для мобильных устройств, смартфонов, имеющих камеры заднего вида и работающих на мобильной операционной системе «android», которая широко распространена в настоящее время. Это значительно увеличило область практического применения данной технологии РП, начиная с рекламы, размещенной на плакатах по аналогии с qr-кодом, и заканчивая защитой документальной информации ограниченного доступа и защитой запчастей к различному оборудованию от контрафакта.

Основные этапы обработки речевой подписи на смартфонах следующие: **распознавание, масштабирование, фильтрация, синтез в волновую форму и озвучивание**. Рассмотрим процесс распознавания РП, предлагаемый для мобильных устройств, более подробно. Описание других этапов рассмотрено в работах [1, 3 – 5].

Распознавание речевой подписи на мобильных устройствах

Алгоритм распознавания речевой подписи для смартфонов представлен на рис. 3.

Блок схема процесса анализа видеопотока и выделение в нем речевой подписи

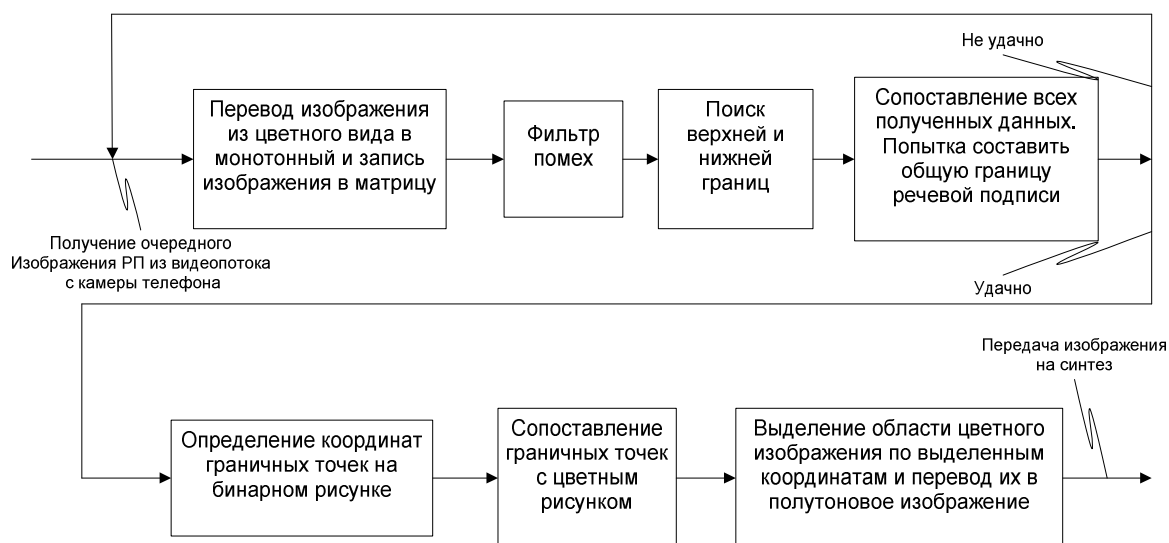


Рис. 3. Схема алгоритма распознавания речевой подписи

Как видно из рис. 3, алгоритм распознавания изображения РП состоит из следующих действий.

1. Получение цветного изображения с РП из видеопотока с камеры смартфона.
2. Перевод полученного цветного изображения в бинарный вид и сохранение этого изображения в виде матрицы числовых интервалов, характеризующих яркости цвета по вертикали.
3. Возможный пропуск данной матрицы через фильтры помех и сглаживания. Одним из таких фильтров является медианный фильтр сглаживания.

4. Поиск в этой матрице границ речевой подписи и анализ ее характерных признаков.
 5. В случае удачного выделения границ речевой подписи определяются граничные точки.
 6. Сопоставление граничных точек на бинарном рисунке цветному рисунку.
 7. Копирование участка цветной картинку.
 8. Перевод нового изображения в монохромный вид для дальнейшего синтеза.
- Для приведенного списка действий необходимо дать следующие комментарии.

Комментарий к пункту 2. Переход цветного изображения к полутонному и бинарному при равномерном освещении осуществляется по следующему алгоритму. Так как каждый пиксель цветного изображения имеет три компоненты яркости: R(красный), G(зеленый), B(синий), и каждая компонента меняется в пределах от 0 до 255, то бинарным изображением будем считать картинку, у которой яркость каждого пикселя вычисляется по формуле:

$$\text{если } \left\lfloor \frac{(R+G+B)}{3} \right\rfloor \geq 128, \text{ тогда } I = 255, \text{ иначе } I = 0. \quad (1)$$

Комментарий к пункту 4. Задача синхронизации – распознавание границ изображения РП (рис. 4) – являлась краеугольной для реализации технологии РП в мобильных устройствах. Под характерными признаками границ речевой подписи подразумевается следующее:

- линии границ четко выделены над и под спектрограммой;
- расстояния между самой нижней линией верхней границы и самой верхней линией нижней границы равны между собой в пределах погрешностей для всего участка речевой подписи.

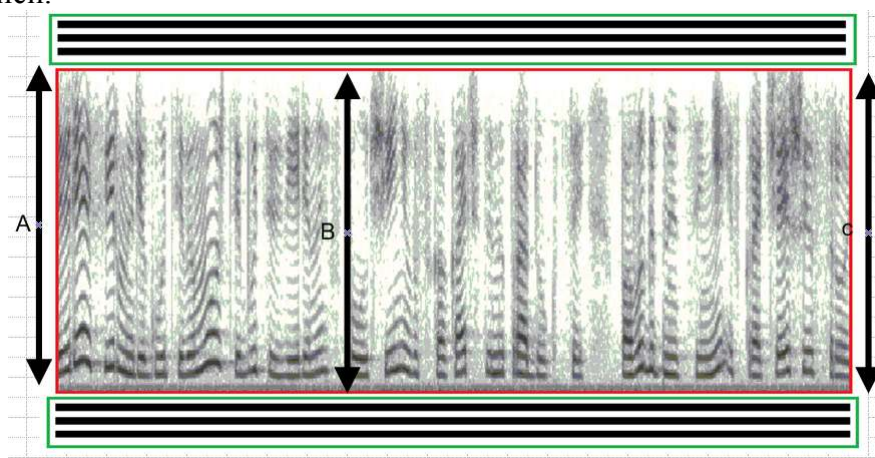


Рис. 4. Изображение речевой подписи (с границами) для использования в мобильных устройствах

Вышеперечисленные действия должны выполняться последовательно для нескольких кадров, и тогда можно считать, что точная граница речевой подписи найдена.

Сам процесс поиска границ заключается в поочередном проходе по сохраненной матрице, в которой хранятся длины отрезков цветов, и нахождение в этих отрезках последовательности, которую можно принять за границу. В данном контексте под последовательностью можно понимать чередование белых и черных полос приблизительно одинаковой длины (или по другому закону – в зависимости от типа границ). После того как часть последовательности в матрице станет соответствовать определенным признакам границы, начнется процесс проверки на сверяемые компоненты (см. рис. 4, 5 и табл. 1). Такие действия проводятся для всей матрицы. Стоит заметить, что, несмотря

на то, что черные и белые линии могут иметь одинаковую ширину, сравниваются они отдельно друг от друга. Это связано с тем, что при захвате изображения с камеры все белые (или черные) полосы могут изменять свою ширину из-за выпуклости объектива. Размеры полос могут отличаться в несколько раз.

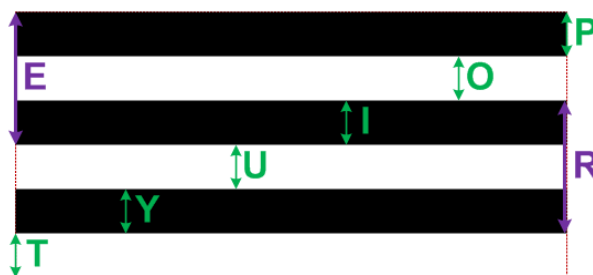


Рис. 5. Границы речевой подписи с выделенными на ней сверяемыми компонентами

Таблица 1. Сверяемые параметры для границы, состоящей из шести одинаковых по ширине, но чередующихся по цвету линий

Сверяемые компоненты	Описание
A, B, C	Данные компоненты характеризуют расстояние между самой нижней линией верхней границы речевой подписи и самой верхней линией нижней границы речевой подписи. Значения данных сверяемых компонент должны совпадать между собой в пределах погрешностей
E, R	Эти компоненты используются для дополнительной проверки распознавания границ, они также должны быть равны между собой
P, I, Y	Данные компоненты характеризуют расстояния между белыми полосками. Значения всех данных расстояний должны быть равны между собой в пределах погрешностей
O, U, T	Эти компоненты характеризуют расстояния между черными полосками. Значения всех данных расстояний должны быть равны между собой в пределах погрешностей

Основным преимуществом такого алгоритма распознавания является то, что он основан, в значительной мере, на сравнении отрезков и не содержит в себе ресурсоемких арифметических операций, которые могли бы усложнить выполнение данного алгоритма на мобильных устройствах малой вычислительной мощности.

Сравнительная оценка зависимости скорости распознавания от вида границ

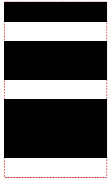
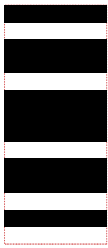
С целью выявления наиболее эффективных границ с точки зрения времени распознавания и занимаемой площади были проведены эксперименты с различными типами границ речевой подписи. С одной стороны, чем время распознавания речевой подписи меньше, тем это лучше для пользователя (потребителя), а с другой стороны, нужно найти такие границы, которые будут занимать меньшую площадь, потому что, чем меньше места занимает граница, тем больше места можно оставить на полезную информацию, т.е. на саму речевую подпись.

К сожалению, эти две задачи являются противоположными друг другу – улучшение результатов одной задачи (например, уменьшение времени распознавания), автоматически приводит к ухудшению результатов другой задачи (увеличению суммарной площади границ). Поэтому необходимо было найти такие границы, которые будут обеспечивать оптимальное соотношение времени распознавания и суммарной площади, которую будет занимать эта граница.

Результаты экспериментов приведены в табл. 2.

Таблица. 2. Сравнительный анализ различных границ РП

№	Вид границы речевой подписи	Описание	Среднее время распознавания	Суммарная площадь общей границы РП в относительных единицах S^*
1		Две чередующихся по цвету полосы одинаковой ширины. Начинается с черной, заканчивается белой	От 20 с до 1–2 мин	4S
2		Четыре чередующихся по цвету полосы одинаковой ширины. Начинается с черной, заканчивается белой	От 12 с до 1 мин	8S
3		Шесть чередующихся по цвету полос одинаковой ширины. Начинается с черной, заканчивается белой	От 3 с до 15 с	12S
4		Восемь чередующихся по цвету полос одинаковой ширины. Начинается с черной, заканчивается белой	От 2 с до 10 с	16S
5		Четыре чередующихся по цвету полосы разной ширины. Третья сверху полоса в два раза больше, чем все остальные. Начинается с черной полосы, заканчивается белой	От 8 с до 50 с	10S
6		Шесть чередующихся по цвету полос разной ширины. Третья сверху полоса в два раза больше, чем все остальные. Начинается с черной полосы, заканчивается белой	От 3 с до 10 с	14S

№	Вид границы речевой подписи	Описание	Среднее время распознавания	Суммарная площадь об- щей границы РП в относи- тельных еди- ницах S*
7		Шесть чередующихся по цвету по- лос разной ширины. Третья сверху полоса в два раза больше, чем белые полосы, а пятая сверху полоса в 3 раза больше, чем белые полосы. Начинается с черной полосы, закан- чивается белой	От 2 с до 6 с	18S
8		Десять чередующихся по цвету по- лос разной ширины. Третья и седь- мая сверху полоса в два раза боль- ше, чем белые полосы, а пятая свер- ху полоса в три раза больше, чем белые полосы. Начинается с черной полосы, заканчивается белой	От 1 с до 4 с	28S

*S – относительная единица площади, которая равна площади одной «единичной» линии на конкретном рисунке.

Результаты проведенных экспериментов показали, что наилучший результат по критерию «времени и площадь» достигается при использовании границ вида № 3 или № 6. Именно они и были выбраны в качестве границ для речевой подписи.

Изображение речевой подписи после ее распознавания может иметь абсолютно произвольные высоту и длину. Это делает затруднительным дальнейшие действия над этим изображением, в частности синтез, вычислительная сложность которого в данном случае будет равняться $(N \cdot M)^2$, где N – длина изображения, M – ширина изображения [3, 5]. Такая сложность не позволит в разумные сроки решать задачу синтеза идентифицированного изображения РП для дальнейшего воспроизведения речи на мобильном устройстве, даже несмотря на то, что современные модели смартфонов имеют 4-х ядерный процессор с тактовой частотой каждого ядра 1,2 ГГц. Поэтому при реализации модуля синтеза речи по изображению РП используется быстрое преобразование Фурье (БПФ), сложность которого возрастает по логарифмической зависимости ($N \cdot \log N$ – для одномерной картинке; $M \cdot N \cdot \log N$ – для двухмерной), а не по квадратичной зависимости, как у обычного дискретного преобразования Фурье. Но для выполнения БПФ требуются определенные начальные условия. В частности, размер изображения РП по вертикали должен являться степенью двойки. Поэтому после задачи распознавания (помимо задач фильтрации и др.) решалась задача масштабирования изображения по оси ординат до 512 пикселей.

Для нахождения значений самих пикселей существует много различных методов, например, метод ближайшего соседа, билинейная интерполяция, бикубическая интерполяция... У всех их есть свои преимущества и недостатки, так, например, метод ближайшего соседа работает в разы быстрее бикубического метода. Но зато качество будет значительно хуже.

Принимая во внимания аспекты разрабатываемой программы, а именно: тот факт, что масштабировать изображение нужно будет всего-навсего один раз, то вполне разумно воспользоваться методом, который даст наилучший результат в плане качества. Под этот метод попадает бикубическая интерполяция, модуль которой также вошел в состав разработанного ПО распознавания и озвучивания РП.

Заключение

В результате выполнения данной работы получены следующие результаты.

1. Разработаны алгоритмы распознавания речевой подписи, лежащие в основе подготовленного программного обеспечения для мобильных устройств считывания (смартфонов) с ОС android.

2. Разработано программное обеспечение для мобильного телефона с камерой небольшой разрешающей способности и процессором средней мощности (~600 МГц). Это дает возможность распознавать и озвучивать аудиомаркеры РП, а затем сравнивать озвученную информацию с ключевыми моментами документа и характеристиками изделия, а также принимать решение о его подлинности и узнавать по голосу автора (собственника) документа или изготовителя изделия.

3. Проведены эксперименты, продемонстрировавшие достаточно высокую работоспособность программы, которая при частоте 30 fps с использованием фильтров шумов и фильтров сглаживания на телефоне с частотой процессора 1,2 ГГц работала без задержек, время обработки одного изображения размером 408×306 было приблизительно равным 0,03 с.

СПИСОК ЛИТЕРАТУРЫ:

1. Дворянкин С.В. Речевая подпись. М.: РИО. 2003, 184 с.
2. Абалмазов Э.И. Компьютерная технология идентификации ценных бумаг // Специальная техника. № 2. 1998. С. 6–9.
3. Алюшин В.М. Цифровая обработка изображений динамических сонограмм для нейтрализации спектральных искажений речевой информации. Диссертация на соискание ученой степени кандидата технических наук. М.: 2014.
4. Алюшин В.М., Дворянкин С.В. Метод реконструкции гармонической структуры спектральных описаний искаженной шумами и помехами речи // Известия института инженерной физики, 2013. Т.2. № 28. С. 57–62.
5. Дворянкин С.В. Взаимосвязь цифры и графики, звука и изображения // Открытые системы. №3. 2000. С. 34–42.

REFERENCES:

1. Dvoryankin S.V. Voicesignature. M.: RIO. 2003. 184 p.
2. Abalmazov E.I. Computer technology securities identification // Special equipment. № 2. 1998. P. 6-9.
3. Alyushin V.M. Digital image processing dynamic spectral sonograms to neutralize the distortion of the speech information. The thesis for the degree of candidate of technical sciences. Moscow, 2014.
4. Alyushin V.M., Dvoryankin S.V. The method of reconstruction of the harmonic structure of the distortion of the spectral noise and interference of speech // Proceedings of the Institute of Engineering Physics, 2013. Vol. 2. №28. P. 57-62.
5. Dvoryankin S.V. Relationship numbers and graphics, sound and image. // Open Systems. № 3. 2000. P. 34-42.