

## АЛГОРИТМ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ ПО ГОЛОСУ

**В.А. ЧАСТИКОВА, З.Я. ТУГУШЕВА, Ф.Р. ГУНАЙ**

*Кубанский государственный технологический университет,  
350072, Российская Федерация, г. Краснодар, ул. Московская, 2,  
электронная почта: zalina.tug@mail.ru*

В статье рассмотрен ряд наиболее популярных алгоритмов идентификации пользователя по голосу: на основе вычитания одного сигнала из другого, алгоритм с применением импульсно - кодовой модуляции и с использованием амплитудно – частотных характеристик. На основе исследованных методов был разработан эффективный комбинированный алгоритм идентификации диктора. В предложенном методе производится разделение звуковой записи на два файла, первый из которых содержит амплитудные значения, а другой - байты служебной информации. Далее проведен сравнительный анализ содержимого первого файла эталонной записи и записи, полученной с микрофона, путем сопоставления нормализованных значений амплитуд. Результат схожести представлен в процентном соотношении.

**Ключевые слова:** идентификация, импульсно-кодовая модуляция, амплитудно-частотная характеристика, сонограмма, амплитуда голоса, эталонная запись.

В статье [1] описан алгоритм нахождения амплитудных характеристик пользователя по голосу. В данной работе предложенный метод исследуется и применяется для дальнейшей реализации алгоритма идентификации пользователя по голосу.

### **1. Алгоритм для считывания необходимой информации с записи**

В данной работе звуковой сигнал записывается в формате .wav, поэтому опишем частный метод считывания данных с .wav файла.

При необходимости непосредственного доступа к буферам со звуковыми данными в программном модуле используется интерфейс низкого уровня, обеспечиваемый несколькими функциями с префиксом имени wave, например, waveInOpen, waveOutOpen, waveOutWrite, waveAddBuffer и т. д.

Данные мультимедиа (звук, видео и т. п.) хранятся в файлах RIFF-формата (ResourceInterchangeFileFormat - формат файла для обмена ресурсами).

В файле формата RIFF содержатся вложенные фрагменты; внешний фрагмент состоит из заголовка и области данных (рисунок 1).

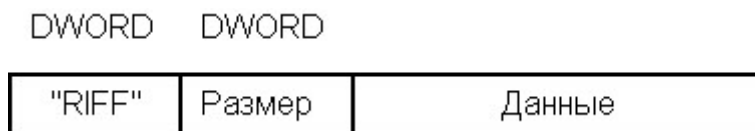


Рисунок 1 - Фрагмент "RIFF".

В первом слове заголовка содержится код FOURCC, в котором идентифицируются данные, находящиеся во фрагменте. Второе двойное слово заголовка - размер области данных в байтах (без учета размера самого заголовка).

Область данных имеет переменную длину, при необходимости она может быть дополнена в конце нулевым байтом до целого числа слов.

Для файла, в котором хранятся звуковые данные (wav-файл), область данных содержит идентификатор данных "WAVE". В файле также может присутствовать фрагмент "LIST" или "INFO", содержащий информацию о правах копирования и другую дополнительную информацию.

Область, обозначенная как "Формат данных", описывает звуковые данные (рисунок 2) [2].

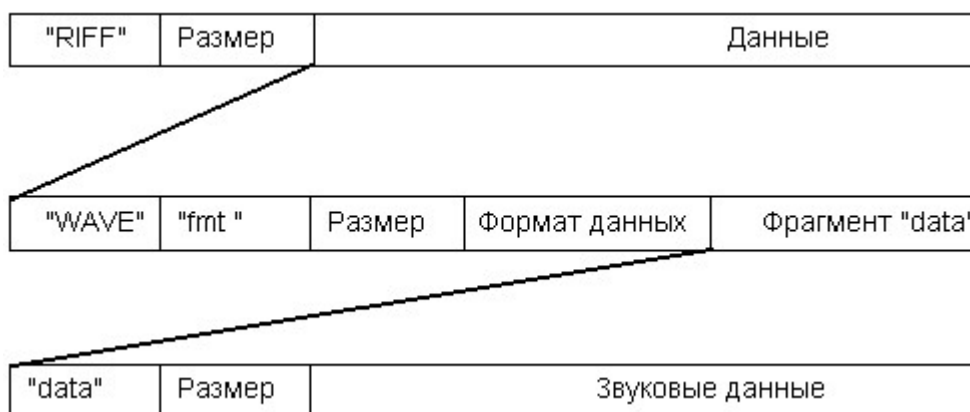


Рисунок 2 - Формат wav-файла.

## 2. Вывод конечного результата

После того как необходимая для идентификации информация будет найдена, данные сравниваются с эталонным образцом, и выявляется степень схожести записи, полученной с микрофона, с записью, находящейся в базе.

Сравнить два звуковых файла можно несколькими вариантами.

*Первый вариант.* Сравнение эталонного и полученного сигналов можно производить путем вычитания одного сигнала из другого (это возможно только тогда, когда начало и конец сигналов точно совпадают) по результирующему сигналу, выведенному путем вычитания одного сигнала от другого. Запись предварительно нужно обработать (нормализация, порезка, очистка от шумов) для получения более точного результата.

*Второй вариант.* В данном случае сравниваются сигналы в цифровом виде и записываются не в абсолютном виде (то есть абсолютные значения амплитуд, Pulse - Code Modulation - Импульсно-кодовая модуляция (ИКМ)), а в виде относительного изменения значений амплитуд сигнала. Другими словами, сигнал представляется значениями, характеризующими изменение значения каждого отсчета относительно предыдущего. Таким образом, результатом является информация об углах наклона амплитудной огибающей в каждой точке или информация о виде (форме) сигнала. После такой обработки происходит вычитание одного сигнала из другого и усреднение значений амплитуды полученной разницы.

*Третий вариант.* Целью такого сравнения является субъективное нахождение различий в звучании двух сигналов. В данном случае для сравнения записей можно использовать несколько разновидностей спектрального анализа, в каждом из которых, тем не менее, имеется ряд недостатков.

1. Первый заключается в графическом сравнении результирующих амплитудно-частотных характеристик (АЧХ) эталонного и полученного с микрофона сжатых сигналов за определенный промежуток времени. Результирующая АЧХ –это график зафиксированных пиковых значений амплитуд, частотных составляющих сигнала за какой-то промежуток времени. Таким образом, взяв два одинаковых промежутка сравниваемых записей и построив их результирующие АЧХ, по совпадению (не совпадению) графиков АЧХ можно приблизительно оценить уровень схожести. Однако этот метод является статичным; он не учитывает изменения сигналов в динамике, а этот

пункт является достаточно важным, так как часто встречаются случаи, когда результирующие АЧХ сигналов почти схожи, однако звучания сравниваемых промежутков сильно отличаются даже на слух.

2. Второй подход для анализа спектра звуковой записи заключается в сравнении сонограмм сигналов (сонограмма - это диаграмма, на которой по оси абсцисс откладывается время, по оси ординат - частота, а амплитуда соответствующей частотной составляющей отмечается интенсивностью цвета в данной точке графика). Сонограмма является наиболее информативной характеристикой и позволяет рассматривать при сравнении изменение сигналов в динамике. Однако этот метод, по сравнению с предыдущим, является более графическим, то есть, если при сравнении статических АЧХ можно оценить "на глаз" различность графиков, то с помощью сонограммы это сделать достаточно трудно, так как сравниваются не кривые графиков, а интенсивность цветов на диаграммах. У сонограмм есть еще один недостаток. Он состоит в размытости сигнала во времени, подразумевающим неточное совпадение (отставание либо опережение) спектральной картины с реальным спектральным составом сигнала в каждый момент времени. Конечно, появление этого эффекта можно в какой-то мере предотвратить с помощью анализа с перекрывающимися окнами в несколько проходов, но объем вычислений при этом сильно увеличивается, и достичь одновременно высоких спектрального и временного разрешений не удастся.

3. Третий метод — это уточненный предыдущий. Его суть состоит в построении АЧХ для каждого окна быстрого преобразования Фурье. Однако такая задача имеет те же проблемы, что и предыдущий метод, и, кроме того, сравнивать графически неудобно, даже если представить всю спектральную картину сигнала в трехмерном виде. [3]

Очевидно, что идеального метода сравнения сигналов не существует. Поэтому в каждом конкретном случае используют наиболее подходящий по точности и удобству метод сравнения из соображений целесообразности.

В данной работе выбран способ сравнения эталонного звукового файла и входного, полученного с микрофона, сигнала на основе результирующих амплитудно-частотных характеристик. Выбор данного метода обусловлен тем, что он является наиболее эффективным для сравнения сигналов, содержащих именно голос, так как у каждого человека индивидуальные АЧХ, и подобный способ дает более точные результаты.

На первом этапе работы была получена графическая АЧХ шаблонного звука и входного сигнала для наглядности работы программного модуля.

Следующая задача - получение всех характеристик сравниваемых звуковых файлов для нахождения результата их схожести. Для ее решения был разработан программный комплекс, позволяющий считывать с записанных и сохраненных файлов формата .wav их характеристики в два файла формата .dat. В одном файле хранятся амплитудные значения сигналов (рисунок 3), в другом - байты служебной информации (рисунок 4).

```

data1.dat  ▸ ×
000018f0  3C FF 37 FF 39 FF 35 FF  34 FF 30 FF 30 FF 2D FF  <.7.9.5.4.0.0.--
00001900  2D FF 29 FF 29 FF 24 FF  24 FF 1F FF 1E FF 1B FF  -.).).$.$.
00001910  1B FF 16 FF 17 FF 13 FF  13 FF 10 FF 0F FF 0C FF  .....
00001920  0D FF 08 FF 08 FF 05 FF  04 FF 01 FF 02 FF FE FE  .....
00001930  FF FE FC FE FB FE F8 FE  F8 FE F6 FE F7 FE F5 FE  .....
00001940  F5 FE F1 FE F4 FE F0 FE  F1 FE F0 FE F1 FE F1 FE  .....
00001950  F0 FE F0 FE F1 FE F1 FE  F1 FE F1 FE F0 FE F2 FE  .....
00001960  F2 FE F4 FE F4 FE F5 FE  F5 FE F7 FE F7 FE FA FE  .....
00001970  FA FE FC FE FC FE FF FE  FF FE 02 FF 02 FF 04 FF  .....
00001980  05 FF 09 FF 09 FF 0F FF  0D FF 13 FF 12 FF 17 FF  .....
00001990  16 FF 1C FF 1B FF 21 FF  21 FF 25 FF 26 FF 2B FF  .....!.!.%.&.+
000019a0  2A FF 30 FF 30 FF 35 FF  35 FF 39 FF 39 FF 3F FF  *.0.0.5.5.9.9.?.
000019b0  3F FF 44 FF 45 FF 4A FF  4A FF 50 FF 4F FF 56 FF  ?.D.E.J.J.P.O.V.
000019c0  56 FF 5B FF 5B FF 5F FF  60 FF 65 FF 65 FF 6D FF  V.[.[._.`.e.e.m.
000019d0  6E FF 73 FF 73 FF 74 FF  74 FF 7A FF 7A FF 8C FF  n.s.s.t.t.z.z...
000019e0  8C FF A6 FF A6 FF BA FF  BA FF C3 FF C3 FF C5 FF  .....
000019f0  C5 FF C6 FF C6 FF CB FF  CB FF CF FF D1 FF D3 FF  .....
00001a00  D4 FF D8 FF D8 FF DB FF  DC FF E0 FF E0 FF E5 FF  .....
00001a10  E5 FF E8 FF E9 FF ED FF  ED FF F0 FF F1 FF F4 FF  .....
00001a20  F3 FF F6 FF F6 FF FA FF  FA FF FD FF FC FF FE FF  .....
00001a30  FF FF FF FF 00 00 01 00  01 00 03 00 02 00 03 00  .....
00001a40  03 00 04 00 04 00 04 00  03 00 02 00 03 00 01 00  .....
00001a50  01 00 01 00 00 00 FF FF  FF FF FF FF FE FF FD FF  .....
00001a60  FC FF FA FF FA FF F8 FF  F8 FF F5 FF F6 FF F3 FF  .....
00001a70  F3 FF F1 FF F1 FF EF FF  EF FF EC FF ED FF E8 FF  .....
00001a80  E8 FF E5 FF E4 FF E2 FF  E1 FF DF FF DF FF DB FF  .....
00001a90  DB FF D8 FF D8 FF D5 FF  D5 FF D1 FF D3 FF CE FF  .....
00001aa0  CE FF CA FF CA FF C7 FF  C7 FF C1 FF C2 FF BE FF  .....
00001ab0  BE FF B9 FF BA FF B6 FF  B6 FF B3 FF B2 FF AD FF  .....
00001ac0  AE FF A9 FF A9 FF A3 FF  A4 FF 9F FF A0 FF 9B FF  .....
    
```

Рисунок 3 – Значения амплитуд голоса.

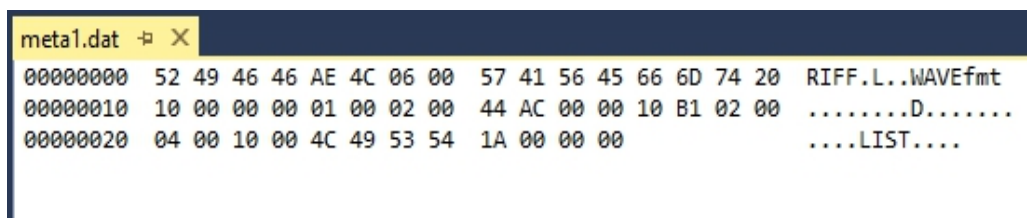


Рисунок 4 – Байты служебной информации.

В данном случае для хранения информации был выбран файл формата .dat так как данные, хранящиеся в нем, могут непосредственно открываться приложениями, и пользоваться такими файлами могут различные программы, но файлы .dat не предназначены для открытия вручную, однако содержимое можно просмотреть с помощью обычного текстового редактора.

После занесения всех звуковых характеристик эталонного образца из базы данных и сигнала, полученного в реальном режиме времени, в отдельные файлы формата .dat, было проведено их сравнение, и выведен результат их схожести в процентном соотношении (рисунок 5).

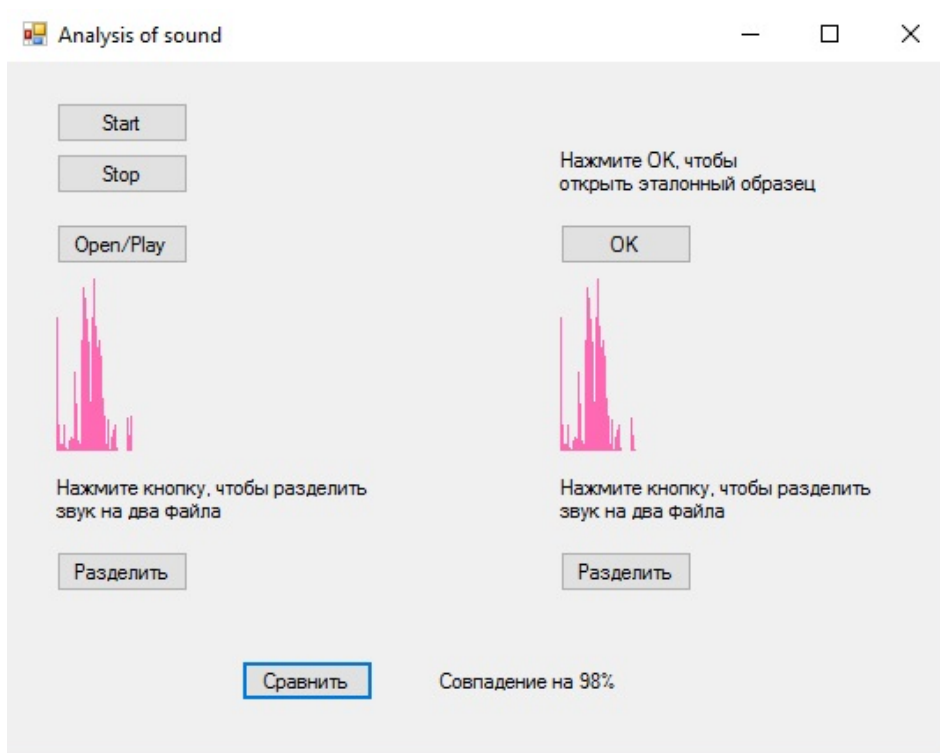


Рисунок 5 — Результат работы приложения.

## ЛИТЕРАТУРА

1. В.А. Частикова, З.Я. Тугушева, Ф.Р. Гунай. Алгоритм нахождения амплитудной характеристики голоса // Научные труды КубГТУ. Краснодар, – 2016 - №2. – [Электронный ресурс] – Режим доступа: URL: <http://ntk.kubstu.ru/file/833>.

2. А. Фролов, Г. Фролов. Мультимедия для Windows, Том 15 , М.: Диалог – МИФИ, 1994, 284 стр.

3. А. Радзишевский. «Заметки на полях» (FAQ). Часто задаваемые вопросы (и ответы) на околозвуковые темы технического характера, 2014. – [Электронный ресурс] – Режим доступа: URL: <http://websound.ru/faq.htm>.

## REFERENCES

1. V.A. Chastikova, Z.Ya. Tugusheva, F.R. Gunay. Algoritm nakhozhdeniya amplitudnoy kharakteristiki golosa // Nauchnye trudy KubGTU. Krasnodar, – 2016 - №2. – [Elektronnyy resurs] – Rezhim dostupa: URL: <http://ntk.kubstu.ru/file/833>.

2. A. Frolov, G. Frolov. Multimediya dlya Windows, Tom 15 , M.: Dialog – MIFI, 1994, 284 str.

3. A. Radzishovski. «Zametki na polyakh» (FAQ). Chasto zadavaemye voprosy (i otvety) na okolozvukovye temy tekhnicheskogo kharaktera, 2014. – [Elektronnyy resurs] – Rezhim dostupa: URL: <http://websound.ru/faq.htm>.

*ALGORITHM OF USER'S VOICE IDENTIFICATION***V.A. CHASTIKOVA, Z.YA. TUGUSHEVA, F.R. GUNAY**

*Kuban State Technological University,  
2, Moskovskaya st., Krasnodar, Russian Federation, 350072;  
e-mail: zalina.tug@mail.ru*

The article describes some of the most popular user identification algorithms voice: based on subtracting one signal from another, algorithm with pulse - code modulation and using amplitude - frequency characteristics. On the basis of the studied methods was developed effective combined algorithm speaker identification. In the proposed method, the separation is performed by two sound recording files, the first of them contains the amplitude values and another - overhead bytes. Next, a comparative analysis of the standard recording and content of the first file received from a microphone, by comparing the normalized amplitude values. Similarity results presented in percentage.

**Key words:** identification, pulse-code modulation (PCM), frequency response, sonogram, the amplitude of the voice, reference record.