

АЛГОРИТМ НАХОЖДЕНИЯ АМПЛИТУДНОЙ ХАРАКТЕРИСТИКИ ГОЛОСА

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 настоящий момент существует множество средств по обеспечению безопасности на основе различных биометрических параметров человека (отпечаток пальца, радужная оболочка, голос, образ лица и т.д.). Такие системы обладают более высокой надежностью, так как базируются на индивидуальных характеристиках людей. Одним из способов защиты является идентификация пользователя по голосу, которая не требует дорогостоящей аппаратуры, достаточно звуковой платы и микрофона. В данной статье описывается алгоритм, записывающий звуковой сигнал, который поступает с микрофона, находит амплитудную характеристику выбранного файла и выводит на экран график полученных значений амплитуд. Также рассмотрены способы нормализации (пиковая и RMS нормализация) и метод «порезки», разделяющий звуковой сигнал на отдельные кадры для проведения сложных операций над большими объемами данных.

Ключевые слова: биометрические параметры, идентификация, аналоговый сигнал, нормализация звука, амплитудная характеристика, пиковая нормализация, RMS нормализация, «порезка» звука.

Биометрические признаки — это индивидуальные, биологически обусловленные характеристики каждого человека. В мире не существует двух людей с одинаковыми биометрическими признаками. Примеры биометрических методов идентификации: анализ отпечатков пальцев, узора радужной оболочки или сетчатки глаз, геометрической формы рук, лица, голоса и т. д. [1]

Большинство биометрических систем безопасности функционируют следующим образом: в базе данных системы хранится цифровой отпечаток пальца, радужной оболочки глаза, голоса или другие биометрические параметры. Для получения доступа к компьютерной сети нужно ввести информацию о себе в систему с помощью микрофона, сканера отпечатков пальцев или других устройств. Поступившие данные сравниваются с образцом, хранимым в базе данных [2].

В качестве биометрического параметра в данном исследовании был выбран голос, так как распознавание по голосу удобно для пользователей и требует от них минимум усилий. Кроме того, почти все современные компьютеры уже имеют необходимое аппаратное обеспечение, а также широкое распространение средств связи (стационарные и мобильные телефонные сети, Интернет-телефония и т. д.) открывает большие возможности для применения данного идентификатора, следовательно, распознавание по голосу является идеальным способом для телекоммуникационных приложений.

Идентификация по голосу основана на анализе индивидуальных характеристик речи человека, зависящих от анатомических особенностей (размер и форма горла и рта, строение голосовых связок) и приобретенными привычками (громкость, манера, скорость речи) [3].

Голос подвержен значительным изменениям под воздействием разных факторов (состояния здоровья или эмоционального состояния). Например, человек с простудным заболеванием может испытывать трудности при пользовании данной системой; также на качество идентификации могут влиять внешние условия (например, посторонние шумы). Если для передачи голосовой информации пользуются линиями связи, помехи в них тоже способны затруднить распознавание пользователя. Существует также возможность воспроизведения звукозаписи с магнитофона. Но большинство проблем, связанных с распознаванием личности по голосу, легко решаются, и данная технология остается наиболее практичным решением для большинства сетевых приложений.

Задача распознавания личности по голосу сводится к выделению и классифицированию голосовых характеристик и соответствующей реакцией на человеческую речь из входного звукового потока. При этом чаще всего выделяют две подзадачи: идентификация и верификация.

Идентификация – процесс определения личности по образцу голоса путём сравнения данного образца с эталонным, сохранёнными в базе данных. Верификация – процесс, при котором с помощью сравнения, представленного

образца с хранимым в базе данных шаблоном, проверяется идентичность. Результат системы может быть положительным (подтверждение личности) или отрицательным (не совпадение голосовых параметров с шаблоном в базе данных) [4].

Целью данной работы является исследование и реализация метода автоматической идентификации личности по голосу.

Для достижения поставленной цели необходимо выполнить следующие задачи:

- 1) Создать приложение, записывающее и сохраняющее звуковую запись, для дальнейшей работы с ней.
- 2) Обработать полученные начальные значения амплитуд для последующей корректной работы с ними.
- 3) Разработать метод, считывающий необходимую информацию со звукового файла для последующего сравнения с эталонным образцом.
- 4) Вывести результат схожести эталонной записи и записи, полученной с микрофона, в процентном соотношении.

В данной статье описан метод получения амплитудной характеристики и способы нормализации этих значений. Рассмотрим подробнее соответствующие пункты.

Приложение, записывающее и сохраняющее звуковую запись

Аналоговые сигналы описываются непрерывной (или кусочно-непрерывной) функцией $x_a(t)$, причем функция и аргумент t могут принимать любые значения на некоторых интервалах $x'_a \leq x \leq x''_a, t' \leq t''$. Вид аналогового сигнала, изменяющегося во времени по закону $x_a(t) = A * e^{-a*t}$, где $A = 1, a > 0$, показан на рисунке 1. Другой пример аналогового сигнала, изменяющегося во времени по закону $x_a(t) = U_m * \sin(2 * \pi * f * t)$, представлен на рисунке 2.

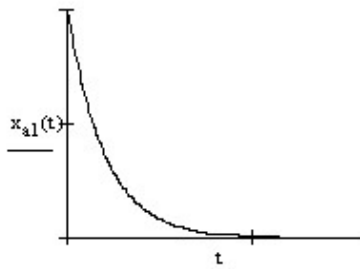


Рисунок 1 — Аналоговый сигнала, изменяющийся во времени по закону $x_a(t) = A * e^{-a*t}$,

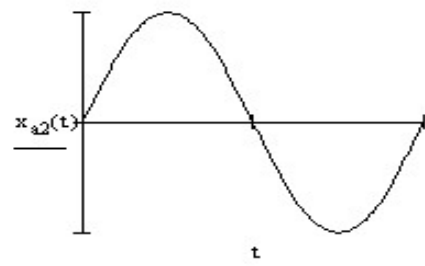


Рисунок 2 — Аналоговый сигнал, изменяющийся во времени по закону $x_a(t) = U_m * \sin(2*\pi*f*t)$

После окончания записи аналогового сигнала он сохраняется в указанной папке. Далее находится амплитудная характеристика полученной записи для дальнейших преобразований.

Амплитудная характеристика - зависимость амплитуды сигнала на выходе устройства от амплитуды сигнала на его входе.

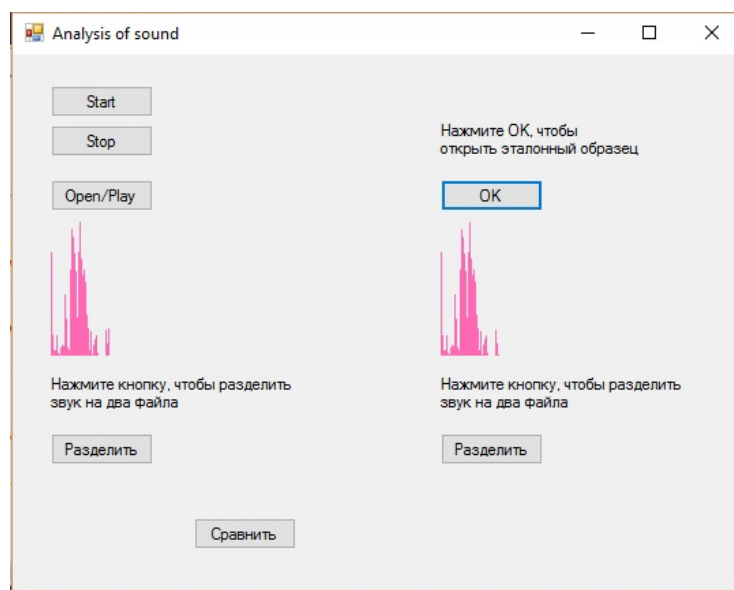


Рисунок 3 — Амплитудная характеристика звукового файла

Нормализация звукового сигнала.

Нормализация звука — процесс выравнивания частотных характеристик при звукозаписи на магнитный носитель.

Способы нормализации

Пиковая нормализация

Автоматизированный способ нормализации, при котором изменяется уровень звукового сигнала путем поднятия до максимально возможного значения для цифрового звука без появления искажений. В данном случае система находит самый громкий пик и по нему выстраивает общий уровень. Этот способ позволяет полностью исключить клиппинг (искажение звукового сигнала, вызванное ограничением амплитуды), но если в звуковом файле есть хотя бы один сильно выделяющийся пик, то при нормализации по его уровню звуковой сигнал останется достаточно тихим, несмотря на то, что звук, являющийся ориентиром, был довольно громким. Для устранения этой проблемы достаточно срезать сильно выделяющийся пик на нужное значение. Величина звука при данном способе обычно измеряется в процентах.

RMS (Root Mean Square) нормализация

Данная нормализация происходит по среднеквадратичному значению уровня звука в файле и абсолютно не похожа на пиковую нормализацию. В рассматриваемом способе величина звука измеряется в *децибелах*. Этот способ лучше всего подходит для человеческого уха и является самым эффективным для поднятия громкости звукового файла, но при большой громкости возможен клиппинг. Для полного избавления от клиппинга, рекомендуется нормализация звука до значения в 89 децибел, но для некоторых слушателей такая громкость может казаться слишком тихой. Также следует учитывать, что если у звуковых файлов разные динамические диапазоны, то на слух они могут звучать с разной громкостью даже с одинаковыми значениями RMS [5].

«Порезка» звукового сигнала.

Даже в звуке с минимально достаточной частотой дискретизации (16 кГц) размер уникальных характеристик для секундного образца звука огромен — 16000 значений амплитуд — и любые сложные операции над такими объемами данных являются невозможными. Кроме того, возникают проблемы со сравнением объектов с разным количеством уникальных черт.

В связи с этим используют метод «порезки» звукового сигнала на отдельные кадры одной длины с половинным перекрытием. Перекрытие является необходимой частью данного процесса, так как возможно искажение звука в случае слишком близкого расположения кадров. Наиболее практичная длина кадра равна 128 мс, так как такая длина является компромиссным вариантом между точностью (длинные кадры) и скоростью (короткие кадры). Оставшуюся речь, не занимающую полный кадр, можно заполнить нулями или отбросить по своему усмотрению [6].

Таким образом, получены искомые звуковые данные в виде амплитудной характеристики для последующего использования их в разработке алгоритма для идентификации пользователя по голосу.

ЛИТЕРАТУРА

1. Борзенко А. Биометрические системы распознавания внешности // Инфраструктурные технологии. ВУТЕ. Россия. - №11 (52). - 2002. – [Электронный ресурс] – Режим доступа: URL: <http://www.bytemag.ru/articles/detail.php?ID=8520>
2. Евангели А. Технологии биоидентификации и биометрический рынок. Виртуальный компьютерный музей. - [Электронный ресурс] – Режим доступа: URL: http://www.computer-museum.ru/technlgy/bio_id.htm
3. Идентификация по голосу. - BioLink.Биометрические истемы. - [Электронный ресурс] – Режим доступа: URL: <http://www.biolink.ru/technology/voice.php>
4. Первушин Е. А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование. - № 3 (24), 2011.- с. 1-54
5. Нормализация звука – Википедия. Свободная энциклопедия: [Электронный ресурс] – Режим доступа: URL: https://ru.wikipedia.org/wiki/Нормализация_звуча
6. Лашко В. Кто там? – Идентификация человека по голосу. Публикации. -[Электронный ресурс]– Режим доступа: URL:<http://habrahabr.ru/post/144491/>

REFERENCES

1. Borzenko A. Biometricheskie sistemy raspoznavaniya vneshnosti // Infrastrukturnye tekhnologii. BYTE. Rossiya. - №11 (52). - 2002. – [Elektronnyy resurs] – Rezhim dostupa: URL: <http://www.bytemag.ru/articles/detail.php?ID=8520>
2. Evangeli A. Tekhnologii bioidentifikatsii i biometricheskiy rynok. Virtualnyy kompyuternyy muzey. - [Elektronnyy resurs] – Rezhim dostupa: URL: http://www.computer-museum.ru/technlgy/bio_id.htm
3. Identifikatsiya po golosu. - BioLink.Biometricheskie istemy. - [Elektronnyy resurs] – Rezhim dostupa: URL: <http://www.biolink.ru/technology/voice.php>
4. Pervushin E. A. Obzor osnovnykh metodov raspoznavaniya diktorov // Matematicheskie struktury i modelirovanie. - № 3 (24), 2011.- s. 1-54
5. Normalizatsiya zvuka – VikipediYa. Svobodnaya entsiklopediya: [Elektronnyy resurs] – Rezhim dostupa: URL: https://ru.wikipedia.org/wiki/Normalizatsiya_zvuka
6. Lashko V. Kto tam? – Identifikatsiya cheloveka po golosu. Publikatsii. - [Elektronnyy resurs]– Rezhim dostupa: URL:<http://habrahabr.ru/post/144491/>

*THE ALGORITHM FOR FINDING THE AMPLITUDE CHARACTERISTIC VOICE***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*

At the moment, there are many means of security based on different biometric parameters of the person (fingerprint, iris, voice, image of the face, etc.). Such systems are more reliable because they are based on the individual characteristics of person. One of methods of protection is to identify the user's voice, which does not require expensive equipment, enough sound card and microphone. This article describes the algorithm of recorded sound signal that comes from the microphone, is amplitude of the selected file and displays a graph of the obtained amplitude values. It is also considered ways of normalization (peak and RMS normalization) and the method of "cutting", which separates the audio signal into separate frames for carrying out complex operations on large amounts of data.

Key words: biometrics, identification, analog signal, audio normalization, amplitude response, peak normalization, the RMS normalization, "cutting" sound.