Информационные процессы, Том 20, № 1, 2020, стр. 10–21

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

Разработка системы аутентификации с использованием верификации диктора по голосу[1]

В.М. Антонова[*,**,***], **К.А. Балакин**[*], **Н.А. Гречишкина**[**], **Н.А. Кузнецов**[**,***]

[*] *Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия*
[**] *Институт радиотехники и электроники им. В.А. Котельникова РАН, Москва, Россия*
[***] *Московский физико-технический институт (национальный исследовательский университет), Московская область, г. Долгопрудный, Россия*

Поступила в редколлегию 28.02.2020

Аннотация—Количество цифровых информационных данных растет с каждым годом. Для доступа к ним используют различные системы идентификации и верификации. Большинство из них предполагают наличие специального ключа (например, пароля), который легко может быть потерян или украден. Использование биометрических данных для доступа исключает эту проблему. В данной работе приведено описание системы идентификации личности по голосу, разработанной в среде MatLab. Помимо функции распознавания, существует возможность устранения шумов и пауз, возникающих в ходе записи.

КЛЮЧЕВЫЕ СЛОВА: Информационная безопасность, аутентификация, обработка звукового сигнала, распознавание речи.

1. ВВЕДЕНИЕ

Первостепенной задачей специалиста по информационной безопасности является разработка такой системы, где вероятность хищения или утечки данных сведена к минимуму. С каждым годом все больше информации, в том числе и составляющей государственную тайну, цифровизируется. Для контроля доступа к важным сведениям применяют различные системы верификации и идентификации. Наиболее популярным подходом является использование персональных паролей и различных идентификаторов личности – логинов, паспортных данных и т. д. Популярность данного метода обусловлена тем, что он не требует наличия дополнительного оборудования, а зачастую и подключения к интернет-сети.

Вероятность хищения, потери или подделки электронного персонального ключа слишком велика. Компании HYPR провела двухлетнее исследование, в ходе которого было опрошенооколо 500 штатных сотрудников США и Канады. Целью было выяснение того, как люди обращаются с личными идентификаторами. Результаты весьма плачевны: 35% пользователей хранят свои пароли в блокнотах, файлах Excel, на стикерах и т. п. Около 78% забывают пароль в течение 90 дней после его создания [1].

Для повышения уровня надежности системперсонализации все чаще используют системы считывания и обработки биометрических данных человека, сложность подделки которых многократно выше, а вероятность потери и вовсе равна нулю. К биометрическим данным относят:

– Анатомические характеристики
– Отпечатки пальцев рук и кистей
– Изображение сетчатки глаза

[1] Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект No 19-07-00525 А).

– Образцы голоса

Каждый параметр уникален и может быть использован в качестве идентификатора пользователя в системе. Методы считывания вышеперечисленных биометрических характеристик имеют как преимущества, так и недостатки. Анатомические характеристики человека (рост, вес, походка и т.д.) являются наиболее легко воспроизводимыми параметрами, а их измерение невозможно для систем удаленного доступа. Считывание отпечатков пальцев рук и кистей предполагает непосредственный контакт с аппаратурой, что существенно ограничивает область применения. Анализ изображения сетчатки глаза может быть опасен для здоровья в случае малейшего сбоя аппаратуры.

Наибольшим потенциалом обладают системы верификации, основанные на считывании образцов голоса. Во-первых, для подобного рода систем отсутствует привязка к внешним условиям. Во-вторых, получить образец голоса можно как при непосредственном контакте со считывающим устройством, так и дистанционно. В-третьих, единственным требованием к оборудованию является наличие микрофона, что не является проблемой для большинства современных устройств. Внедрение систем верификации возможно не только в системах доступа к помещениям или данным. Они могут использоваться для разблокировки рабочих и персональных компьютеров, смартфонов, планшетов и даже автомобилей.

Большинство современных систем безопасности, основанных на использовании биометрических данных человека, работают следующим образом: с помощью различного рода сканеров и считывателей в систему вводятся данные, которые впоследствии сравниваются с неким образцом – заранее внесенными “правильными” данными. Наиболее эффективны системы, использующие смешанный подход. Например, для входа в систему необходимо ввести логин и пароль, а после произнести секретную фразу или ввести отпечаток пальца.

По-другому устроены системы распознавания речи диктора. Их работа делится на два этапа: обучение и распознавание. На первом этапе на базе особенностей речи конкретного человека формируетсямодель, являющаяся идентификатором конкретного пользователя. На втором этапе сформированная модель сравнивается с вводимой в систему фразой [2].

В данной работе в среде MatLab была создана система безопасности с использованием верификации диктора по голосу. Главной задачей была разработка такого продукта, который бы позволил производить все операции в едином окне, а также обеспечивал бы вариативность использования. Интерфейс приложения представлен на Рис. 1.

Работу системы можно представить в виде двух схем (Рис. 2 и Рис. 3):

На первом этапе в систему поступает сигнал, распространяющийся в виде волн в воздушной среде (см. Рис. 4). Как было сказано ранее, приемником в данном случае выступает микрофон, по умолчанию установленный на большинстве современных устройств. Задачей данного устройства является преобразование механических колебаний в электрические. Обилие всевозможных микрофонов (электростатические, электродинамические, угольные и т.д.) объединяет одно – принцип действия. Важнейшим элементом является высокочувствительная мембрана, совершающая колебания при воздействии на нее звуковых волн. С помощью преобразующего элемента механические колебания мембраны интерпретируются в электрический сигнал [3].

На этапе первичной обработки сигнал подвергается следующим манипуляциям [3, 4]:

1. Усиление уровня сигнала.
   Для снижения нагрузки на вычислительные ресурсы ЭВМ, уровень сигнала необходимо усилить.
2. Высокочастотная фильтрация.
   Последующее применение фильтра высоких частот позволяет скомпенсировать снижение интенсивности сигнала, которое составляет порядка 20dB/dec. Как итог – увеличение ве-

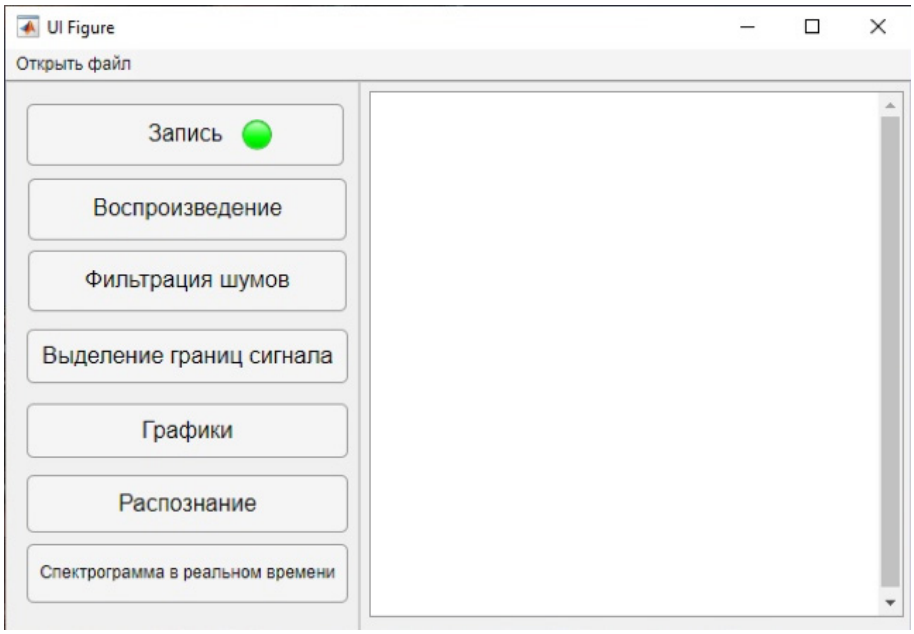


Рис. 1. Интерефейс приложения.

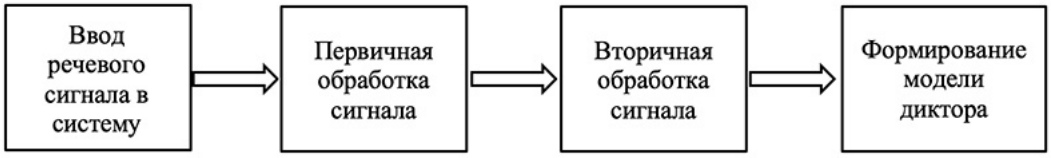


Рис. 2. Этап обучения системы распознавания диктора.

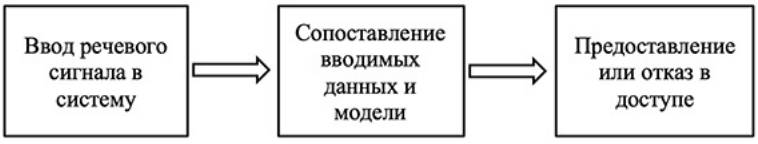


Рис. 3. Этап распознавания диктора.

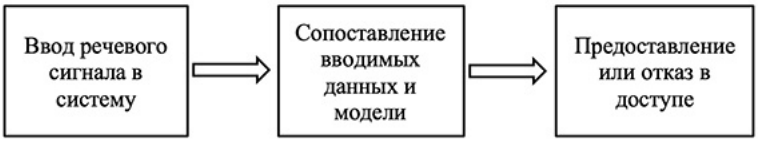


Рис. 4. Исходный сигнал, поступающий в систему.

личины отношения высоких частот к более низким.

$$Y_t = X_t - k \times X_{t-1}, \tag{1}$$

где t – время; k – коэффициент предискажений, численно равный 0.97.

3. Фильтрация шумов.
Шум – беспорядочные звуковые колебания, характеризующиеся случайными изменениями амплитуды и частоты. Источником его является двигающийся воздух. Необработанный сигнал поступает в устройство в виде суперпозиции беспорядочных колебаний и требуемого сигнала. Чистота речевого сигнала влияет на скорость распознавания и дальнейшей обработки. Поэтому качество звука необходимо максимально высокое. Для снижения влияния шума и получения четких спектральных характеристик применяют различного рода фильтры. Например:

$$x_t = (x_t - 0.9x_{t-1}) \left(0.54 - 0.46 \cos\left[(t-6) \times \frac{2\pi}{180^\circ}\right]\right), \tag{2}$$

где t – время; x_t – набор дискретных значений звукового сигнала.
По завершении трёх вышеперечисленных пунктов получаем преобразованный сигнал (см. Рис. 5):

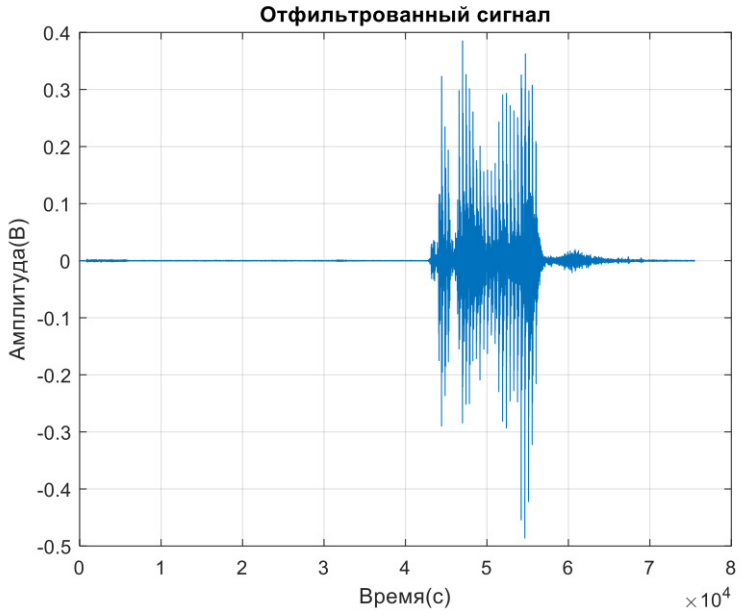


Рис. 5. Сигнал, пропущенный через первый ряд фильтров.

4. Выделение границ речевого сигнала.
На данном этапе из поступающего сигнала выделяют значимую информацию, отбрасывая все паузы и случайные звуки. Наиболее эффективным подходом является анализ плотности распределений значений отсчетов пауз. Первые 200 мсек записи речевого сигнала, как правило, являются паузой, значения отчетов которой – случайные величины, распределенные по нормальному закону. На основе полученной информации о плотности распределения отсчетов паузы выполняется выделение границ речи. Далее система анализирует участки с минимальными амплитудами – паузами, отбрасывая их. Отфильтрованный таким образом сигнал представлен на Рис. 6.

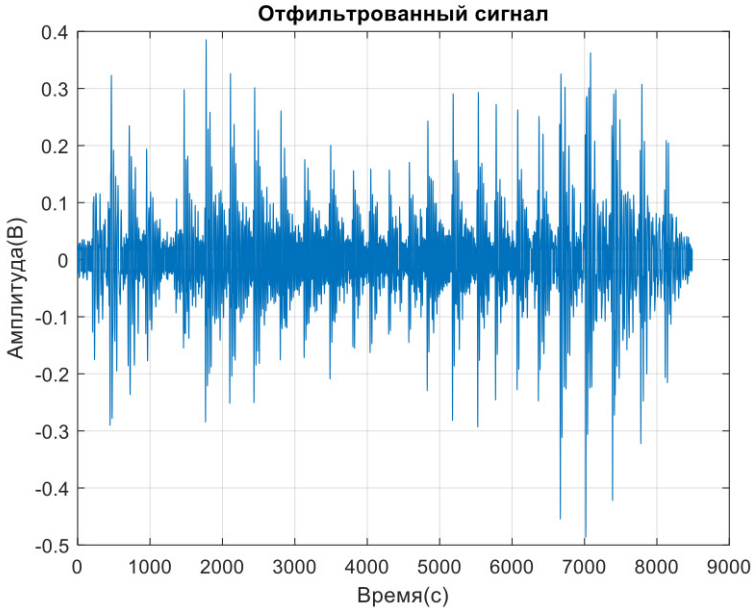


Рис. 6. Сигнал после первичной обработки.

Этап вторичной обработки сигнала состоит из следующих пунктов [4, 5]:

1. Деление на фреймы.
   Поступающие звуковые сигналы слишком сложны, чтобы сразу их обрабатывать. Так, звук, имеющий минимальную достаточную частоту дискретизации – около 16 кГц, имеет около 16 значений амплитуд для миллисекундного образца. Существует еще одна проблема: на данный момент не понятно, как проводить сравнение объектов, имеющих разное количество уникальных характеристик.
   Решением является разбиение сигнала на большое число составляющих определенного размера. Одинаковый размер подсигналов и последующее усреднение результатов вычислений позволят сразу получить необходимое число признаков для последующей классификации.
   Поступающий после этапа первичной обработки сигнал делится на фреймы – сегменты фиксированной длины. Обычно длина их составляет порядка 10–60 мсек. Длину и точку отсчета подбирают таким образом, чтобы получившиеся сегменты не шли строго друг за другом, а перекрывались. Делается это для предотвращения возможной потери информации на границах раздела. Уровень перекрытия при этом обычно составляет 20–60%. Чем меньше перекрытие, тем меньшей размерностью в итоге будет обладать вектор свойств, характеризующий данный участок. Таким образом, получается постоянной длины окно, движущееся вдоль исходной функции.
   Рассмотрим, как происходит наложение фреймов на примере первых двух фрагментов (см. Рис. 7). Первый фрейм отмечен синим цветом, второй желтым. Перекрытие в данном случае составляет 50%.
   Как показали эксперименты, наилучший результат достигается при длине фреймов в 60 мсек.
2. Дискретное преобразование и оконная обработка фреймов [2, 3, 7].
   Пройдя через два этапа предварительной обработки, формируются данные, содержащие информацию об амплитуде, частоте и форме огибающей речевого сигнала.
   Стоит отметить, что использование в качестве численной характеристики фрейма среднего квадрата всех его значений (RootMeanSquare) не применяется, так как в данном случае эта метрика не является информативной [6].

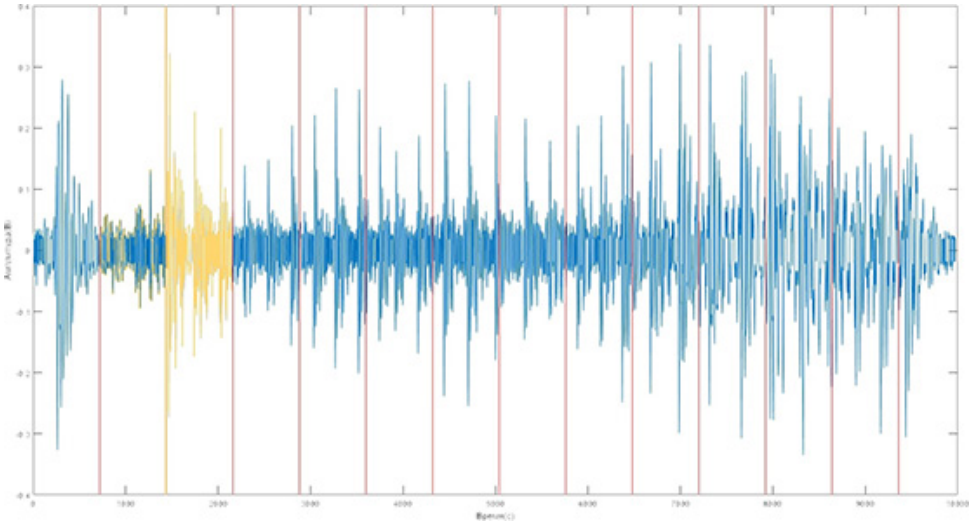


Рис. 7. Сегментирование сигнала с наложением.

Одним из возможных методов получения устойчивых критериев, характеризующих диктора, является метод вычисления спектра мощности и коэффициентов Фурье через дискретное преобразование Фурье для каждого фрейма.

Для снижения граничных эффектов, неизбежно возникающих при делении сигнала на фреймы, используют различные весовые (“оконные”) функции: прямоугольное окно, окно Кайзера, окно Блэкмана, окно Хэмминга и т.д.

Оконная функция Win(n) определяется следующей формулой:

$$Win(n) = (1 - \alpha) - \left(\alpha \times \cos\left(\frac{2 \times \pi \times n}{N - 1}\right)\right), \tag{3}$$

где n – порядковый номер фрейма, для которого вычисляется новое значение амплитуды; N – длина фрейма; α – коэффициент, характеризующий конкретную оконную функцию.

Наиболее часто используемой функцией является окно Хэмминга, задающееся формулой

$$Win(n) = 0.53836 - 0.46164 \times \cos\left(\frac{2 \times \pi \times n}{N - 1}\right), \tag{4}$$

где n – порядковый номер фрейма, для которого вычисляется новое значение амплитуды, N – длина фрейма.

Пусть $X_i(k)$ – результат преобразования i-ого фрейма:

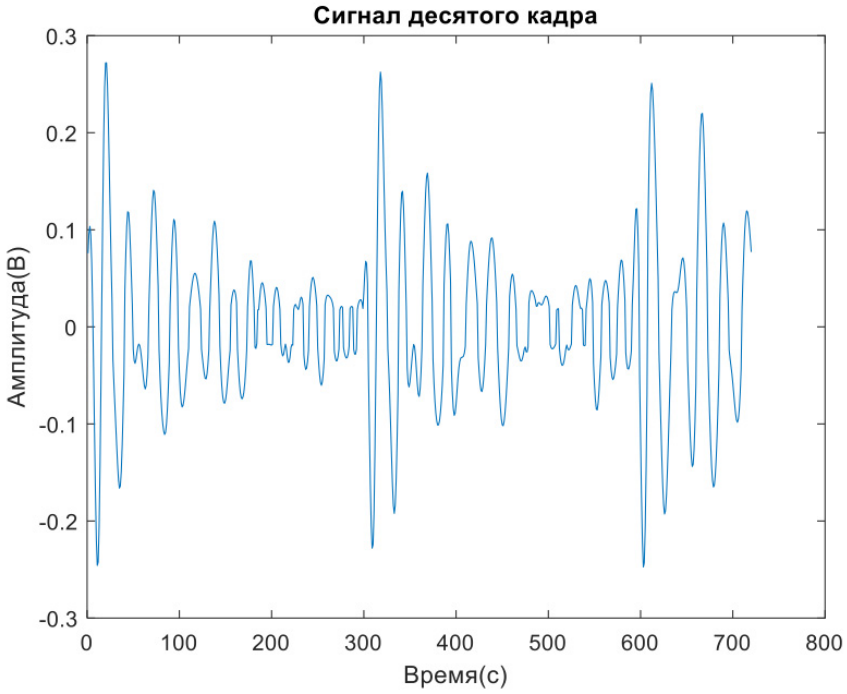
$$X_i(k) = \sum_{n=0}^{N-1} x_i \times Win(n) \times e^{-\frac{2\pi i}{N} kn}, \qquad k \in (1, \ldots, N); \tag{5}$$

где N – длина фрейма; $Win(n)$ – оконная функция; x_n – амплитуда n-го сигнала; x_k – число комплексных амплитуд синусоидальных сигналов, слагающих исходный сигнал.

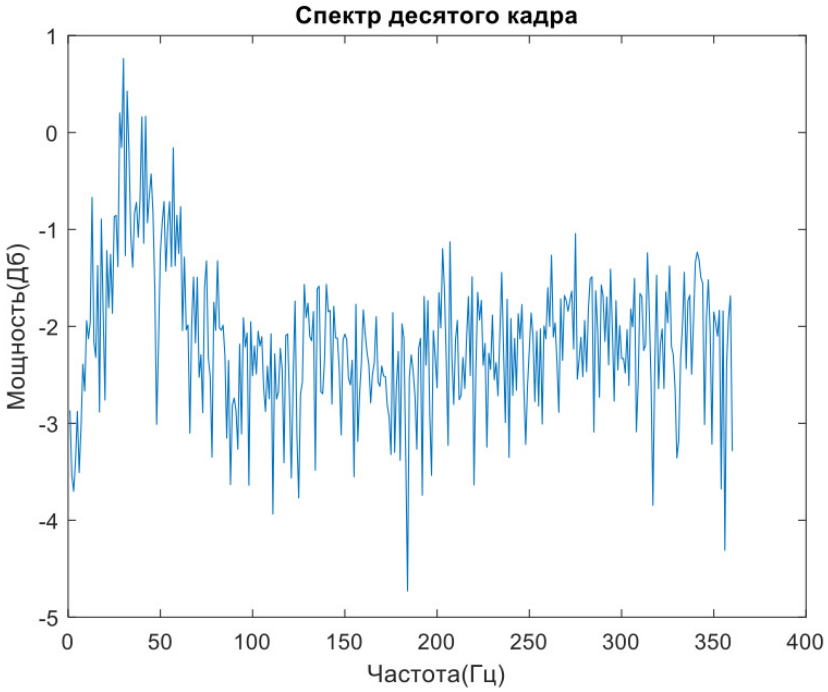
Для наглядности рассмотрим в увеличенном масштабе любой фрейм, например 10, и применим для него быстрое преобразование Фурье (см. Рис. 8).

Для оценки спектральной плотности сигнала, вычислим периодограмму:

$$P_i(k) = \frac{1}{N} |X_i(k)|^2 \tag{6}$$



(a) Десятый фрейм звукового сигнала



(b) Десятый фрейм после Дискретного преобразования и наложения окна Хэмминга

Рис. 8. Обработка десятого кадра.

Здесь берется абсолютное значение комплексного числа, полученного в результате преобразования Фурье, а также исключается вторая половина коэффициентов Фурье из-за их

симметрии относительно центра. Существуют алгоритмы быстрого преобразования(FFT – Fast Fourier Transformation), дающие возможность весьма быстрой обработки.

3. Переход к шкале Мел [2, 3, 5, 8, 9].
   Данный этап позволит получить наборустойчивых признаков для каждого диктора. Метод кепстрального представления, т.е. обратного Фурье-преобразования от логарифмического амплитудного спектра в масштабе частот Мел, схожим с масштабом восприятия сигнала человеком, был разработан для такого описания речевого сигнала, которое было бы устойчиво к индивидуальным особенностям дикторов [6, 8].
   Общая формула, связывающая частоты в Мел и линейной шкалах, выражается как:

$$M(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right), \tag{7}$$

   где f – частота в Гц, M – частота в мелах.
   Обратный переход осуществляется как

$$M^{-1}(m) = 700 \times \left(e^{\frac{m}{1125}} - 1\right). \tag{8}$$

Чтобы разложить спектр входящего сигнала по Mel-шкале, пропустим его через ряд фильтров. Каждый Mel-фильтр – это треугольная оконная функция, применяемая к вычисленной выше периодограмме,позволяющая суммировать энергию на определённом диапазоне частот. Эта сумма и есть Mel-коэффициент. Число фильтров лежит в пределах от 20 до 40, обычно используется 26 фильтров. Эмпирическим путем было выяснено, что наилучший результат достигается при использовании 36 Мел-фильтров.

Устойчивыми признаками, характеризующими диктора, будут являться так называемые Мел-частотные кепстральные коэффициенты (MFCC – Mel-frequency cepstral coefficients). MFCC – это вектор, обычно состоящий из 13 вещественных чисел. Кепстральные коэффициенты формируют пространство, в котором будет производиться распознавание диктора. Как правило используют от 10 до 30 коэффициентов. Часто используются первые и вторые разности по времени кепстральных коэффициентов, что втрое увеличивает размерность пространства принятия решений, но улучшает эффективность распознавания диктора [2, 7, 8].

Алгоритм нахождения MFCC выглядит следующим образом [8]:

1. Преобразование нижней и верхних частот спектра в Мел-шкалу для получения рабочего диапазона.
2. Введение дополнительных точек в Мел-диапазон.
3. Перевод значений из Мел-шкалы в значения в СИ.
4. Каждой точке периодограммы соответствует определенная частота, вычисляемая по формуле:

$$f(i) = \frac{2(i+1)}{F \times n}, \tag{9}$$

   где F – частота дискретизации.
5. Расчёт фильтров.
   Mel-фильтр – это вектор, значения которого отличны от нуля лишь на определенных промежутках:

$$H_m(k) = f(x) = \begin{cases} 0, & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \leqslant k \leqslant f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \leqslant k \leqslant f(m+1) \\ 0, & k > f(m+1). \end{cases} \tag{10}$$

Результат – трехмерная матрица коэффициентов, представленная на Рис. 9.

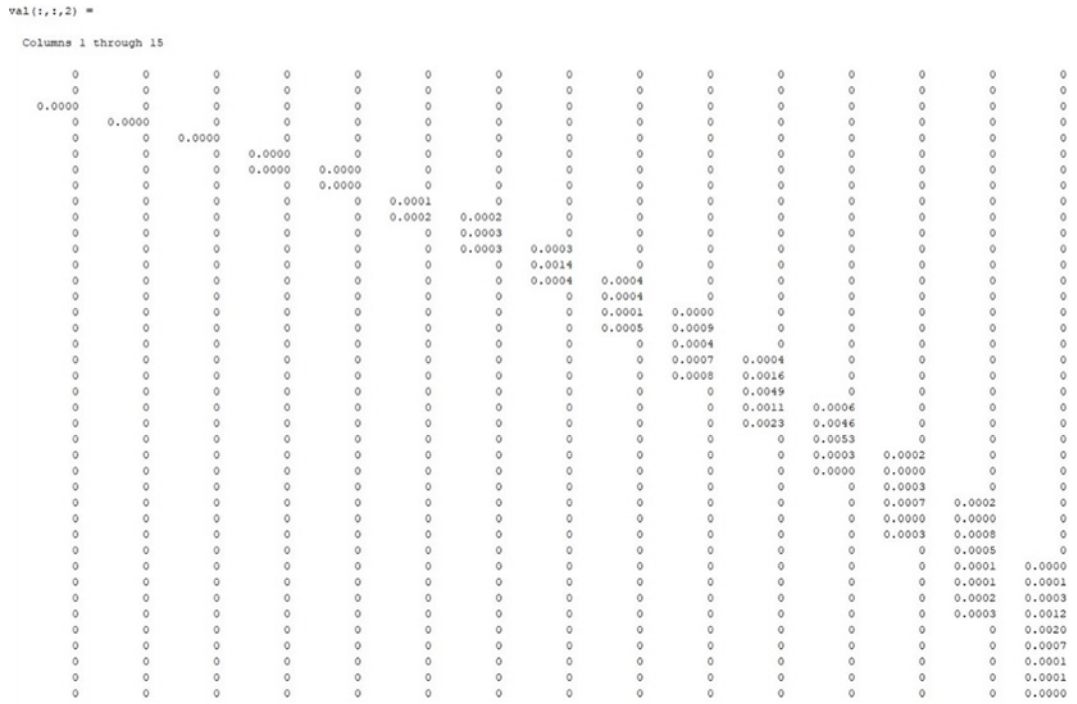


Рис. 9. Трехмерная матрица коэффициентов.

6. Вычисление энергии для каждого полученного окна:

$$E(m) = \ln\left(\sum_{k=0}^{N-1} P_i(k) * H_m(k)\right), \qquad m \in [0; M). \tag{11}$$

Применение данной формулы позволяет получить численное распределение энергии полученных окон. Результат представлен на Рис. 10.

7. Дискретное косинусное преобразование полученных величин для вычисления MFCC:

$$c(n) = \sum_{m=0}^{M-1} E(m) * \cos\left(\frac{\pi n(m+\frac{1}{2})}{M}\right), \qquad n \in [0; M). \tag{12}$$

На Рис. 11 представлен первичный ветор прихнаков, характеризующий диктора.

8. Применение метода нормализации коэффициентов
Вычисляем кепстральное среднее (см. Рис. 12) с целью снижения динамических изменений, вызванных каналом передачи. Данное значение приближенно описывает спектральные характеристики канала передачи, например, микрофона

$$\overline{c_t} = c_t - \frac{1}{T}\sum_{i=0}^{T-1} c_i. \tag{13}$$

9. Расширение вектора признаков
Дополним вектор признаков временными изменениями – первыми и вторыми производными кепстральных коэффициентов. Возможные сверточные искажения сигнала обычно медленно изменяются во времени и аддитивны в кепстральной области. Такое дополнение позволит снизить их влияние.

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	-15.0202	-13.5424	-13.6563	-12.9312	-12.1392	-13.1375	-13.5404	-13.0311	-10.5808	-11.7469	-14.3453
4	0	-11.6030	-14.3524	-15.2644	-12.9296	-11.6240	-14.5778	-14.4623	-15.1489	-12.2436	-11.9906	-12.4667
5	0	-10.5395	-11.8298	-13.4930	-12.1984	-12.7886	-12.5497	-12.6687	-14.9609	-11.6765	-11.2100	-11.5306
6	0	-11.2454	-13.2115	-14.1016	-13.7513	-11.8314	-12.9521	-13.6924	-11.2102	-10.9098	-11.7583	-11.0654
7	0	-12.7215	-11.5140	-10.8611	-12.7215	-13.0471	-12.4183	-12.5264	-12.5254	-10.4169	-10.5511	-10.2193
8	0	-10.7470	-12.5333	-12.5612	-14.1726	-11.7502	-12.6581	-12.4088	-12.1912	-12.4803	-8.1317	-9.1542
9	0	-9.1781	-12.6647	-9.9359	-15.1959	-12.6365	-11.8590	-12.2584	-15.1915	-10.9784	-9.5678	-7.0790
10	0	-7.9173	-12.5587	-10.2510	-11.3307	-11.5806	-10.6870	-10.0645	-10.2769	-9.4190	-7.5581	-7.9889
11	0	-7.9749	-13.6827	-9.5139	-11.3140	-12.2635	-17.8249	-11.4850	-8.9572	-9.5753	-7.7995	-6.9937
12	0	-7.4632	-10.2736	-8.3778	-8.7291	-10.4157	-9.6832	-9.3569	-9.6326	-9.3150	-6.7545	-5.5131
13	0	-6.5710	-10.4069	-10.1073	-11.0583	-8.7742	-9.5674	-12.0486	-10.7064	-7.8108	-6.9777	-5.8952
14	0	-7.0574	-11.3984	-7.2897	-10.9933	-10.3352	-10.3250	-9.6242	-15.4793	-7.7301	-7.7218	-9.2841
15	0	-7.8794	-9.5645	-10.4284	-7.9889	-10.7006	-10.4157	-8.4471	-9.8270	-9.7653	-8.0585	-8.9123
16	0	-8.8069	-9.8958	-10.4737	-10.6281	-8.0726	-9.2759	-10.3450	-9.7870	-7.5174	-8.5147	-7.4340
17	0	-6.5805	-13.4573	-8.9341	-9.7485	-10.9463	-9.2998	-9.4483	-8.0851	-8.6208	-7.0790	-9.3318
18	0	-7.7437	-13.0800	-10.8176	-8.7893	-9.7942	-7.9464	-8.0580	-12.6669	-10.6140	-9.2863	-7.2081
19	0	-6.8446	-9.5240	-8.0113	-10.5879	-8.2015	-11.2895	-9.5254	-9.7299	-7.7189	-9.7000	-8.3647
20	0	-6.0494	-9.9085	-10.5811	-10.9018	-9.5826	-12.0681	-9.2045	-10.1115	-10.0819	-8.2171	-8.5939
21	0	-5.3236	-8.5454	-12.6469	-8.8869	-10.4833	-10.7947	-9.2777	-10.4838	-10.0442	-8.4442	-9.3793
22	0	-6.4028	-9.9787	-7.2454	-8.8981	-10.4177	-8.6082	-9.5272	-8.7517	-8.7980	-8.9869	-7.3048
23	0	-4.9728	-9.0223	-10.0650	-8.6414	-8.4203	-8.2289	-8.0089	-9.1223	-9.3600	-9.1722	-8.0948
24	0	-5.2432	-11.5196	-8.6637	-7.9572	-10.1704	-11.8031	-10.9936	-10.5996	-8.8657	-7.0247	-6.1857
25	0	-7.5999	-8.0262	-7.5397	-11.2473	-8.5326	-8.9749	-9.1530	-7.7153	-7.6223	-7.9195	-7.3041
26	0	-10.9112	-8.3283	-9.3644	-9.7986	-8.2467	-10.5811	-7.7692	-8.9591	-8.3812	-7.1300	-5.3867
27	0	-8.2413	-8.9480	-6.4960	-8.4939	-8.5450	-8.0550	-8.3824	-7.5301	-11.9720	-5.8375	-7.0561
28	0	-7.0456	-8.8952	-6.4868	-10.8316	-8.2357	-9.2172	-6.7578	-8.5515	-6.4180	-5.1329	-5.2764
29	0	-11.1994	-8.8710	-6.9956	-7.2411	-7.4372	-8.7158	-7.4414	-7.4978	-6.4262	-6.0253	-8.0563
30	0	-6.8881	-9.7552	-4.7319	-9.3903	-6.7584	-7.6262	-8.2117	-8.0652	-8.2405	-6.2405	-7.1032
31	0	-7.5678	-8.4043	-5.7341	-9.3160	-9.5152	-8.7206	-5.7065	-7.2198	-5.7460	-5.3807	-7.3374
32	0	-8.8301	-9.4789	-7.1157	-6.1459	-5.5023	-6.5341	-5.6347	-8.5444	-4.8614	-6.0804	-7.0470
33	0	-8.3378	-7.7952	-6.1486	-7.9469	-7.5111	-8.6090	-6.1126	-5.7285	-7.8124	-7.7030	-8.0458
34	0	-7.6156	-9.4902	-10.5820	-8.0304	-6.7256	-6.2827	-5.7490	-8.7026	-6.6609	-6.6253	-8.0888
35	0	-6.4819	-9.3455	-6.2325	-6.8415	-5.4731	-6.3844	-6.6025	-8.1317	-6.4580	-8.2746	-7.6712
36	0	-6.2131	-8.5706	-6.0583	-7.1072	-9.8218	-5.9017	-6.3338	-5.3616	-6.7753	-7.6585	-6.7251
37	0	-7.0950	-7.3150	-8.6323	-8.4435	-6.5025	-6.6733	-7.0776	-7.4401	-7.5522	-9.5509	-7.4699
38	0	-8.8204	-9.2317	-6.6579	-7.2867	-7.6822	-6.9058	-7.8629	-5.9769	-7.1422	-7.8365	-5.9678
39	0	-8.3899	-8.4358	-7.3857	-8.7663	-6.4600	-7.2775	-7.9813	-6.9545	-6.3232	-6.4190	-6.4268
40	0	-10.1838	-7.1969	-5.9324	-8.9550	-8.1028	-7.1677	-5.4383	-6.5036	-6.3408	-6.5236	-5.7607
41	0	-8.5130	-8.7992	-6.1620	-9.5431	-5.9039	-8.2575	-7.6233	-8.3091	-6.0319	-7.6687	-6.0006
42	0	-9.4666	-7.9119	-7.2041	-11.0043	-7.6588	-7.0195	-6.6112	-8.1526	-8.7556	-5.4967	-6.2171
43	0	-10.0912	-8.1263	-5.9218	-6.3886	-7.4085	-7.0476	-5.9827	-9.3691	-9.9797	-6.0629	-5.7947
44	0	-9.2022	-7.7610	-7.3190	-7.0308	-5.8434	-6.3655	-8.0822	-7.5346	-6.0399	-8.0397	-6.5297
45	0	-8.1394	-7.1215	-7.0921	-9.7241	-7.1950	-7.6098	-7.2931	-10.7913	-8.0635	-6.7113	-8.5227
46	0	-8.6596	-9.5453	-9.8011	-6.6199	-7.4398	-7.3259	-8.2789	-7.8258	-7.2303	-8.4079	-6.9575
47	0	-10.6510	-8.2760	-8.5698	-7.2687	-6.7122	-8.0382	-11.2722	-7.8826	-7.2408	-6.2943	-6.9665
48	0	-9.8530	-6.8459	-9.0768	-8.7236	-8.8564	-7.8508	-7.4698	-8.5650	-8.7347	-8.4209	-8.5468
49	0	-9.4144	-7.3612	-7.2210	-10.1887	-8.6744	-9.5905	-6.8484	-7.7680	-8.9443	-7.6716	-7.1048
50	0	-8.8013	-11.0147	-10.1423	-9.4092	-9.6297	-8.5156	-7.0580	-11.3131	-8.2293	-7.6317	-9.0539
51	0	-9.6818	-9.7853	-8.3977	-9.4378	-8.1532	-7.8000	-8.9568	-11.4017	-9.8356	-9.3686	-7.3894
52	0	-8.5196	-9.1986	-7.9774	-7.2582	-9.9431	-9.1532	-8.0548	-10.5732	-10.2059	-8.1087	-8.7936
53	0	-8.0116	-7.6239	-10.2928	-9.3738	-7.2663	-8.3558	-6.5633	-11.2711	-8.6134	-9.0227	-6.1793
54	0	-8.1930	-10.0685	-7.9973	-8.7949	-7.8255	-7.8495	-7.6943	-8.7337	-8.5576	-7.9396	-9.1138
55	0	-8.0919	-10.2457	-10.2005	-8.1324	-9.2997	-9.2446	-9.4219	-10.7129	-10.4891	-10.1245	-6.7256
56	0	-7.9658	-9.4665	-7.4455	-9.4178	-7.4257	-8.0842	-7.5862	-10.3895	-8.2769	-7.9032	-9.3047
57	0	-8.6122	-9.8144	-8.1801	-11.3306	-9.7601	-9.9582	-8.3449	-10.1436	-11.0804	-8.1167	-7.1136
58	0	-9.9267	-11.1814	-9.3705	-8.7245	-8.1434	-8.4496	-9.1684	-12.5122	-10.7577	-10.4662	-12.3677
59	0	-10.3783	-12.0352	-10.5994	-11.8226	-9.3761	-9.2316	-8.9600	-11.1159	-8.3199	-8.9844	-8.8439

Рис. 10. Трехмерная матрица коэффициентов.

	1
1	0
2	-20.5345
3	54.1028
4	15.5738
5	44.0466
6	26.3001
7	51.5132
8	25.9835
9	45.4778
10	14.9942
11	36.9058
12	12.5522

Рис. 11. Первичный вектор признаков.

Полученный вектор признаков, по сути являющийся материалом для построения будущей модели диктора, обрабатывается с помощью алгоритм адинамической трансформации временной шкалы (DTW – dynamic time warping), целью которых является нахождение оптимальных соответствий для временных масштабов слов [10].

	1
1	-25.5763
2	-46.1108
3	28.5265
4	-10.0025
5	18.4703
6	0.7239
7	25.9369
8	0.4072
9	19.9015
10	-10.5821
11	11.3295
12	-13.0240

Рис. 12. Нормализованный вектор признаков.

DTW – алгоритм вычисляет оптимальную последовательность деформации времени между двумя временными рядами и расстояния векторами признаков

$$d(v, w) = \sum_{i=1}^{N} |v_i - w_i|. \tag{14}$$

По завершении работы алгоритма, каждого диктора будет характеризовать некоторое число (набор чисел). На основании выявленных закономерностей и происходит процесс идентификации.

Вышеописанный алгоритм уже реализован в среде MatLab. Его синтаксис представлен на Рис. 13.

Syntax

```
dist = dtw(x,y)
[dist,ix,iy] = dtw(x,y)

[__] = dtw(x,y,maxsamp)

[__] = dtw(__,metric)

dtw(__)
```

Рис. 13. Алгоритм DTW в MatLab.

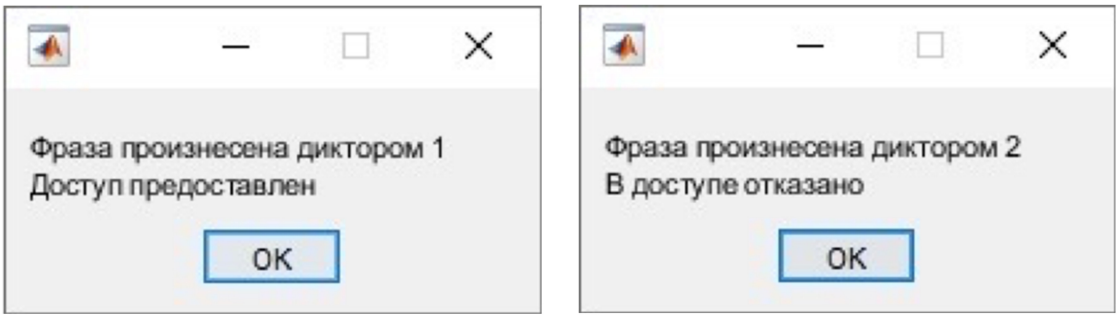


Рис. 14. Итог работы приложения.

Итогом работы приложения является отказ или предоставление доступа к системе (см. Рис. 14).

2. ЗАКЛЮЧЕНИЕ

В ходе выполнения данной работы была успешно создана система верификации диктора по голосу. Система обладает довольно высокой степенью надежности. Для тестирования были выбраны 10 дикторов, образ и манера речи которых кардинально отличались. По итогу 9 участников эксперимента были успешно распознаны. Работа имела экспериментальный характер, поэтому выборки были короткие. При увеличении длины выборки вероятность распознавания может быть существенно увеличена.

СПИСОК ЛИТЕРАТУРЫ

1. Kraft E. New password study by hypr finds 78% of people had to reset a password they forgot in past 90 days.

2. В.Н.Сорокин, В.В.Вьюгин, А.А.Тананыкин, Распознавание личности по голосу: аналитический обзор // *Информационные процессы,* 2012, Том 12, No1, стр. 1–30.

3. В.Н.Сорокин, А.И.Цыплихин, Верификация диктора по спектрально-временным параметрам речевого сигнала // *Информационные процессы,* 2010, Том 10, № 2, стр. 87–104.

4. Ле Н. В., Панченко Д. Предварительная обработка речевых сигналов для системы распознавания речи // Молодой ученый. 2011. - №5. Т.1. - С. 60–75.

5. Rabiner L.R., Juang B.-H. (1993). Fundamentals of speech recognition. PTR Prentice Hall, Englewood Cliffs, N.Y.

6. Young, S., A Review of Large-Vocabulary Continuous Speech Recognition, IEEE Signal. Processing Magazine, pp. 21-60, Sep. 1996

7. Lu X., Dang J. An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. *Speech Communication,* (2007) v.50, N4, pp. 151–320.

8. Furui S. Cepstral analysis techniques for automatic speaker verification. IEEE Tran. Acoust., Speech, Signal Processing, (1981) v. 27, pp. 250–280.

9. Miyajima C., Watanabe H., Kitamura T., Katagiri S. (1999). Discriminative feature extraction - Optimization of Mel-cepstral features using second-order all-pass warping function. Proc. EUROSPEECH, II-779-I-782.

10. S. Salvador, P. Chan Fast DTW: Toward Accurate Dynamic Time Warping in Linear Time and Space. Intelligent Data Analysis 11.5 (2007), pp. 561–580.

Development of an authentication system using voice announcer verification

V.M. Antonova, K.A. Balakin, N.A. Grechishkina, N.A. Kuznetsov

The amount of digital information data is growing every year. Various identification and verification systems are used to access them. Most of them require a special key (for example, a password), which can easily be lost or stolen. Using biometric data for access eliminates this problem. This article describes the voice identification system developed in the MatLab. In addition to the recognition function, it is possible to eliminate noise and pauses that occur during recording.

KEYWORDS: Information security, authentication, audio processing, speech recognition.