



## ВЛИЯНИЕ ПСИХОФИЗИОЛОГИЧЕСКОГО СОСТОЯНИЯ ДИКТОРА НА ПАРАМЕТРЫ ЕГО ГОЛОСА И РЕЗУЛЬТАТЫ БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ ПО РЕЧЕВОМУ ПАРОЛЮ\*

Сулавко А. Е.<sup>1,2</sup>, Еременко А. В.<sup>2</sup>, Борисов Р. В.<sup>3</sup>, Иниватов Д. П.<sup>1</sup>

<sup>1</sup>Омский государственный технический университет, Омск, Россия

<sup>2</sup>Омский государственный университет путей сообщения, Омск, Россия

<sup>3</sup>Сибирский государственный автомобильно-дорожный университет, Омск, Россия

### Аннотация

В работе использовались два метода для вычисления идентификационных характеристик голоса диктора. Один из них основан на прямом преобразовании Фурье, второй — на оконном преобразовании с последующей интеграцией значений каждой гармоники всех окон. Определена информативность данных характеристик. Дана оценка того, каким образом параметры голоса и их информативность изменяются в зависимости от степени алкогольного опьянения человека и в сонном состоянии. Проведен вычислительный эксперимент по оценке надежности распознавания дикторов в пространстве выбранных признаков с помощью функционалов, основанных на формуле гипотез Байеса, мере Пирсона, мере хи-модуль, критериях Джини, Крамера-фон Мизеса, а также перцептронов, обучаемых по ГОСТ Р 52633.5-2011, и сетей квадратичных форм. Дана оценка устойчивости указанных методов и функционалов к изменению психофизиологического состояния диктора в плане робастности получаемых результатов распознавания.

**Ключевые слова:** *распознавание образов, параметры речевого сигнала, голосовой пароль, биометрическая аутентификация, психофизиологическое состояние диктора, состояние алкогольного опьянения.*

**Цитирование:** Сулавко А. Е., Еременко А. В., Борисов Р. В., Иниватов Д. П. Влияние психофизиологического состояния диктора на параметры его голоса и результаты биометрической аутентификации по речевому паролю // Компьютерные инструменты в образовании. 2017. № 4. С. 29–48.

### 1. ВВЕДЕНИЕ

В настоящее время, как никогда, актуальной является проблема защиты конфиденциальной информации от неавторизованных пользователей. Классическим способом решения проблемы является защита данных с помощью паролей, шифрования либо аппаратных ключей. Однако эти способы являются отчуждаемыми от владельца, в случае

\*Работа выполнена при финансовой поддержке РФФИ (грант №15-07-09053).

попадания ключей и паролей третьим лицам защищенная информация будет скомпрометирована. Данную проблему пытаются решить с помощью аутентификации по биометрическим признакам. Большинство существующих коммерческих решений базируется на статических биометрических образах (лица, отпечатков пальцев, узоров радужной оболочки глаза и т. п.). Фундаментальная проблема, лежащая в использовании открытых биометрических образов, заключается в том, что они не являются секретными и могут быть фальсифицированы (путем создания муляжа пальца, использования высококачественной фотографии и т. п.). На сегодняшний день ведутся активные исследования динамических биометрических признаков для решения аналогичных задач, в частности параметров голоса. Если голосовой пароль является одноразовым и генерируется случайным образом непосредственно перед аутентификацией, проблема изготовления «муляжа» биометрического образа частично решается. Однако характеристики голоса являются менее информативными по сравнению со статическими биометрическими признаками и дают больший процент ошибок. Кроме того, они меняются в зависимости от психофизиологического (функционального) и эмоционального состояний диктора. Часто в процессе разговора мы легко можем определить, что знакомый нам человек находится в состоянии алкогольного опьянения или только что проснулся. Настоящая работа направлена на повышение надежности голосовой аутентификации и оценку того, насколько указанные состояния влияют на идентификационные параметры речевых паролей. Об актуальности предлагаемого проекта можно судить по следующим данным. В соответствии с Национальной технологической инициативой (НТИ), к приоритетным научно-техническим задачам относится разработка технологий биометрической идентификации и аутентификации личности. НТИ — это долгосрочная комплексная программа по созданию условий для обеспечения лидерства российских компаний на новых высокотехнологичных рынках, которые будут определять структуру мировой экономики в ближайшие 15–20 лет. В проекте НТИ уделяется внимание созданию рынка устройств, работающих с биомониторингом для обеспечения безопасности, и устройств, позволяющих управлять персональной цифровой собственностью при помощи биокриптографической подписи.

Согласно Доктрине информационной безопасности (ИБ) РФ (от 5 декабря 2016 года), в последние годы возрастают масштабы компьютерной преступности. В соответствии с Доктриной ИБ к основным направлениям обеспечения ИБ в областях науки относятся:

- достижение конкурентоспособности российских информационных технологий и развитие научно-технического потенциала в области обеспечения информационной безопасности;
- создание и внедрение информационных технологий, изначально устойчивых к различным видам воздействия;
- проведение научных исследований и осуществление опытных разработок в целях создания перспективных информационных технологий и средств обеспечения информационной безопасности.

Создание систем голосовой аутентификации по биометрическим образам, устойчивых к изменению состояния диктора, относится к приоритетным задачам научно-технического развития рынка информационной безопасности.

## 2. ПОСТАНОВКА ЗАДАЧИ

Имеется несколько трактовок понятия «состояние человека» [1]. Интерес представляют следующие:

1. Функциональное состояние (ФС), которое характеризует эффективность деятельности или поведения человека и возможность выполнить конкретную работу. Диагностика ФС выполняется на основании результатов измерения психофизиологической информации, а также информации о качестве деятельности субъекта.
2. Психофизиологическое состояние (ПФС) — совокупность свойств человека, отражающих биологические аспекты проявления адаптации к изменяющимся условиям окружающей среды и оцениваемых на основании измерения психофизиологической информации.
3. Эмоциональное состояние (ЭС) — психическое состояние, которое возникает в процессе жизнедеятельности субъекта и определяет не только уровень информационно-энергетического обмена, но и направленность поведения.

Приведенные выше понятия имеют сходства, но и имеют различия. Разные авторы опираются на различные уровни функционирования человека (физиологический, психологический, психофизиологический, эмоциональный). Термин «функциональное состояние» часто употребляется в контексте рассмотрения эргатических систем и когда речь идет о производительности труда. В работе [2] приводится несколько точек зрения относительно смыслового различия понятий ФС и ПФС. Автор не считает понятия ФС и ПФС тождественными, однако подчеркивает, что они взаимосвязаны, и придерживается термина ПФС. Авторы настоящей статьи также придерживаются данного термина. Цель настоящей работы: определить характер влияния ПФС субъекта на биометрические параметры произнесенных им голосовых паролей и вероятность ошибок распознавания диктора с использованием различных мер близости.

Биометрический параметр, характеризующий субъекта, который извлекается из речевого сигнала и представляет собой отдельную физическую величину, будем называть признаком. Надежность процедуры верификации образов определяется вероятностью (или процентом) ошибок 1-ого и 2-ого рода (False Rejection Rate (FRR) и False Acceptance Rate (FAR), соответственно). В настоящей работе критерием эффективности меры близости (надежности принимаемых решений) считается средняя оценка вероятности ошибок 1-ого и 2-ого рода  $((FRR + FAR)/2)$ .

### 3. ФОРМИРОВАНИЕ БАЗЫ ОБРАЗЦОВ ГОЛОСОВЫХ ПАРОЛЕЙ

Для участия в натурном эксперименте по сбору данных голоса было привлечено 86 испытуемых в возрасте от 18 до 35 лет, мужчин и женщин (в равном соотношении), без выраженных заболеваний или неврологических нарушений (предварительно проводилась проверка неврологического статуса, пульса, частоты сердечных сокращений и др.). Эксперимент проводился в начале рабочего дня после полноценного отдыха. Испытуемые последовательно «вводились» в разные ПФС. В каждом ПФС испытуемые не менее 60 раз произносили 8 голосовых паролей, состоящих из 2-х слов («идентифицируйте меня», «разрешите доступ» и другие). Аудиозаписи голосов дискретизированы с размером аудиообразца — 16 бит и частотой дискретизации 8000 Гц. Второй параметр задавался исходя из диапазона частот, занимаемых речевым сигналом (до 4000 Гц), в соответствии с теоремой Котельникова кодирование без потерь для непрерывного сигнала из диапазона до определенной частоты возможно при его дискретизации с удвоенной частотой. Использовались следующие ПФС (где  $s$  — порядковый номер состояния):

1. Исходное состояние (нормальное,  $s = 1$ ). В данном состоянии субъект не подвергался никаким воздействиям.

2. Состояния алкогольного опьянения. Опираясь на Федеральные правила полетов США (91.17: Алкоголь и пилотирование) и методические указания Минздрава (от 03.07.1974 «О судебно-медицинской диагностике смертельных отравлений этиловым алкоголем и допускаемых при этом ошибках»), было выделено 3 условных стадии, каждая из которых определяет конкретное ПФС субъекта:
  - 2.1. Первая стадия (опьянение 1,  $s = 2$ ), при котором содержание алкоголя в крови составляет от 0.2 до 0.3 %. Согласно методическим указаниям Минздрава, в этом состоянии отсутствует существенное влияние алкоголя на организм. Однако возможны скрытые проявления нарушений, которые могут быть обнаружены специальными тестами.
  - 2.2. Вторая стадия (опьянение 2,  $s = 3$ ), при котором содержание алкоголя в крови составляет от 0.3 до 0.5 %. Согласно методическим указаниям Минздрава, такая концентрация оказывает незначительное влияние на организм. Алкоголь в таком количестве может влиять на мыслительный процесс, решения, координацию и концентрацию, а также может увеличивать время реакции. Наблюдается легкая эйфория, расслабление, ощущение радости, снижение предупреждений, понижение сдержанности.
  - 2.3. Третья стадия (опьянение 3,  $s = 4$ ), при котором содержание алкоголя в крови составляет от 0.5 до 1 %. Согласно методическим указаниям Минздрава, концентрация алкоголя в пределах 0.5–1.5 % соответствует легкому опьянению. Количество алкоголя от 0,5 % приводит к статистически значимым изменениям вариабельности сердечного ритма [3, 4]. Согласно используемым источникам, на данной стадии опьянения наблюдаются притупление ощущений, экстравертность, нарушаются способность к рассуждениям, глубина восприятия, периферическое зрение, реакция зрачка на свет.
3. Сонное состояние (сон,  $s = 5$ ). Данное ПФС характеризуется низкой продуктивностью. Для имитации данного состояния испытуемые принимали естественные растительные средства седативного действия (пустырник, мяту, валериану). ЧСС снижалась на 3–5 % по сравнению с нормальным состоянием.

Концентрация алкоголя в крови рассчитывалась по формуле Видмарка:

$$c = A/m \cdot r, \quad (1)$$

где  $c$  — концентрация алкоголя в крови в %,  $A$  — масса выпитого напитка в граммах,  $m$  — масса тела в килограммах,  $r$  — коэффициент Видмарка (0,7 — для мужчин, 0,6 — для женщин).

Таким образом, определено 5 состояний. Все ПФС кроме нормального также назовем измененными.

#### 4. ГОЛОСОВЫЕ ПРИЗНАКИ СУБЪЕКТОВ И ОЦЕНКА ИХ ИНФОРМАТИВНОСТИ

Для статистического анализа сообщений произвольной длины и неконтролируемого уровня громкости стоит задача нормализации по амплитуде и длине образца записи, состоящая в поиске преобразования, в результате которого получался бы вектор фиксированной размерности, сравнимый с аналогичными результатами таких преобразований по величине каждого из элементов результирующего вектора-отпечатка. Предложено два метода такого преобразования, поэтому условно разделим используемые в работе признаки на 2 группы.

Первая строится на интегрировании амплитудного спектра речевого сигнала  $\bar{U}$  в окрестности экстремумов. Предварительно спектр разбивается на некоторое число отрезков, равное количеству признаков. Данный подход взят из работ [5, 6]. В первой из них [5] вычислялось по 60 признаков из парольной фразы целиком, во второй [6] — из каждой гласной фонемы. Однако в работе [6] не учитывались случаи ошибочного сегментирования речевого сигнала (ложное выделение фонемы, ложный пропуск фонемы). Количество ошибок при использовании методов разбиения, основанных на корреляционном анализе речевых сигналов из [6], в большинстве случаев оказалось значительным. Поэтому в настоящем исследовании решено не делить парольную фразу на фрагменты, а обрабатывать сигнал целиком, получая из него вектор  $\bar{a}_1$  значений некоторого количества признаков.

В работе предлагается несколько иной подход для получения второй группы признаков. Массив  $\bar{U}$  может быть представлен как множество интервалов значений  $\bar{U}_k$ , длина каждого из которых равна  $n$  отсчетов сигнала, а их количество  $q$  определяется шириной окна и шагом окна. Каждый интервал  $\bar{U}_k$  — это массив чисел длины  $n$ , к которому применяется быстрое преобразование Фурье (БПФ):

$$\bar{U}_k = (u_0, u_1, \dots, u_i, \dots, u_{n-1}, \dots, u_n) \xrightarrow{\text{БПФ}_k(\bar{U}_k)} FFT_k.$$

В результате действия БПФ получается массив комплексных чисел  $FFT_k$ , для каждого из которых вычисляется амплитудно-частотная характеристика (АЧХ) —  $\bar{A}$ :

$$FFT \xrightarrow{A_i = \sqrt{fft_i \cdot re^2 + fft_i \cdot im^2}} \bar{A},$$

где вектор  $\bar{A}_k$  — это массив амплитуд гармоник, сумма которых равна сумме элементов  $\bar{U}_k$ , а количество амплитуд гармоник в силу свойств БПФ равно  $n/2$ :  $\bar{A}_k = \{A_{k_1}, A_{k_2}, \dots, A_{k_i}, \dots, A_{k_{n/2}}\}$ . Далее вычисляется энергия интервала сигнала для последующей нормализации АЧХ:

$$E_k = \frac{2 \sum_i^{n/2} \sqrt{fft_{k_j} \cdot re^2 + fft_{k_j} \cdot im^2}}{n}$$

Вариативность громкости сигнала определяется силой произнесения, направления относительно микрофона и расстояния от него. Проблема изменчивости амплитуды речевого сигнала решается нормализацией АЧХ интервала сигнала по уровню. Для этого осуществляется деление амплитуды каждой гармоники  $A_j$  интервала  $\bar{U}_k$  на  $A_k$ :

$$\bar{A}_k \xrightarrow{A'_{k_i} = A_{k_i} / E_k} \bar{A}'_k,$$

где элемент  $\bar{A}'_k = \{A'_{k_1}, A'_{k_2}, \dots, A'_{k_i}, \dots, A'_{k_{n/2}}\}$  — нормализованная АЧХ интервала  $k$ . При таком подходе учитывается, что условия записи могут меняться даже на протяжении короткой ключевой фразы, подвергаемой обработке.

Для каждой совокупности  $i$ -х гармоник всех интервалов вычисляются средние арифметические значения  $ma_i$ , являющиеся оценкой момента математического ожидания амплитуд  $i$ -й частоты:

$$ma_i = \frac{2 \sum_{k=1}^q A'_{k_i}}{n}$$

Совокупность оценок этих моментов есть вектор значений признаков второй группы, представляющий собой интегральную оценку частотного наполнения сигнала фиксированной размерности для сообщений различной длины:  $\bar{a}_2 = \{a_1 = ma_1, a_2 = ma_2, \dots, a_i = ma_i, \dots, a_{(n/2)} = ma_{(n/2)}\}$ .

Таким образом, каждый сигнал  $\bar{U}$ , формируемый после однократного произнесения диктором парольной фразы, преобразуется в векторы значений признаков  $\bar{a}_1$  и  $\bar{a}_2$ . Объединенный вектор значений признаков (первой и второй группы) обозначим как  $\bar{a}$ .

Используемые в настоящей работе признаки зависят от амплитудно-частотных характеристик звукового сигнала, поэтому одновременно содержат информацию как о дикторе, так и о произнесенной им парольной фразе. Кроме того, данные признаки могут содержать информацию о состоянии человека. Последнее является гипотезой, проверка которой выходит за рамки настоящей работы. На данный момент достоверно можно утверждать, что ПФС субъекта находит отражение в речевом сигнале и соответственно вносит шумы, влияющие на рассматриваемые биометрические признаки. Данное исследование, прежде всего, направлено на оценку силы и характера этого влияния с точки зрения задачи распознавания диктора по голосу и не преследует целей поиска в этих шумах информационной составляющей, характеризующей состояние субъекта. Однако некоторые полученные результаты говорят в пользу подтверждения высказанной гипотезы.

Под информативностью признака понимается то, насколько хорошо он характеризует распознаваемых субъектов. Чем выше информативность признака, тем с меньшей вероятностью ошибки данный признак позволяет разделить классы образов дикторов. Сумма вероятностей ошибок 1-го и 2-го рода распознавания двух дикторов по некоторому голосовому признаку стремится к площади пересечения подграфиков функций плотностей вероятности значений этого признака, характеризующих соответствующие субъекты. В настоящем исследовании по данным выборки оценивались площади пересечения между всеми имеющимися парами подграфиков функций плотностей вероятности, характеризующих различные субъекты, аналогично тому, как это выполнялось в работе [7]. Данные площади всегда больше нуля и не могут превысить единицу.

Для всех признаков  $a_j$  вычислены площади пересечения  $S_p$  между всеми возможными парами имеющихся подграфиков функций плотностей вероятности  $f_{psi}(a_j)$ , заданных на области значений  $a_j$ , каждая из которых характеризует  $i$ -го диктора, произносящего  $p$ -й пароль, и находящихся в ПФС под номером  $s$ . Информативность признака  $a_j$  зависит от среднего значения вычисляемых площадей и задачи, относительно которой производится оценка. С точки зрения распознавания парольной фразы, произнесенной известным диктором, информативность  $a_j$  можно оценить по формуле:

$$\bar{I}(a_j, s) = 1 - \frac{\sum_{i=1}^N Sp(f_{1si}(a_j), f_{2si}(a_j))}{N} \quad (2)$$

для задачи распознавания диктора по фиксированной фразе целесообразно использовать формулу:

$$\bar{I}(a_j, s, p) = 1 - \frac{2 \sum_{i=1}^N \sum_{l=i+1}^N Sp(f_{psi}(a_j), f_{psi}(a_j))}{N \cdot N - N}, \quad (3)$$

при распознавании диктора по тайной парольной фразе оценка может быть получена по формуле:

$$\bar{I}(a_j, s) = 1 - 2 \frac{\sum_{p=1}^{p\_count} \sum_{t=p+1}^{p\_count} \frac{2 \sum_{i=1}^N \sum_{l=i+1}^N Sp(f_{psi}(a_j), f_{psi}(a_j))}{N \cdot N - N}}{p\_count \cdot p\_count - p\_count}, \quad (4)$$

Формально формула (4) соответствует задаче одновременного распознавания и фразы и диктора, то есть фразы, произнесенной определенным диктором. Отметим, что если бы речь шла о распознавании рассматриваемых ПФС некоторого (известного заранее) субъекта, то для вычисления информативности  $a_j$  можно было воспользоваться формулой:

$$\bar{I}(a_j, p) = 1 - \frac{\sum_{i=1}^N \sum_{s=1}^5 \sum_{l=s+1}^5 Sp(f_{psi}(a_j), f_{psi}(a_j))}{10N}, \quad (5)$$

если о верификации нормального состояния известного диктора – формулой:

$$\bar{I}(a_j, p) = 1 - \frac{\sum_{i=1}^N \sum_{s=2}^5 Sp(f_{psi}(a_j), f_{psi}(a_j))}{4N}, \quad (6)$$

где  $n$  — количество испытуемых,  $p\_count$  — число голосовых паролей.  $\bar{I}(a_j)$  в данном случае является величиной, противоположной взвешенной сумме площадей пересечения подграфиков плотностей вероятности (обозначаемой иногда как  $I(a_j) = 1 - \bar{I}(a_j)$ ), и измеряется от нуля до единицы, что наиболее удобно для целей настоящего исследования, но при необходимости может быть переведена в биты информации по формуле [8]:

$$I_{bit}(a_j) = -\text{Log}_2(I(a_j)), \quad (7)$$

Оба подхода позволяют получить произвольное количество признаков  $\eta_1$  и  $\eta_2$  ( $\eta = \eta_1 + \eta_2$ ). Чем больше признаков используется при верификации образов, тем ниже их средняя информативность, но выше суммарная. Суммарная информативность признаков — это всегда взвешенная сумма частных оценок информативности признаков с учетом их взаимной корреляции. Целесообразно выбрать оптимальное число признаков из каждой группы и использовать их совместно (рис. 1). В результате исследования установлено, что, с точки зрения задачи распознавания диктора (по фиксированной парольной фразе), наиболее информативными признаками первой условной группы  $\bar{a}_1$  являются те, что ассоциированы с частотой ниже 400–450 Гц (наименьшей информативностью обладают признаки, характеризующие высокую частоту). Информативность признаков второй условной группы  $\bar{a}_2$  распределена более равномерно, их рекомендуется использовать все (рис. 1).

Важно то, что признаки  $\bar{a}_1$  и  $\bar{a}_2$  вычисляются различными способами, поэтому корреляция между ними преимущественно слабая. При комплексировании разнородных признаков появляется большее число пар признаков с низким коэффициентом корреляции (рис. 2). Из представленных рисунков видно, что признаки первой группы менее зависимы, чем признаки второй группы, однако и менее информативны.

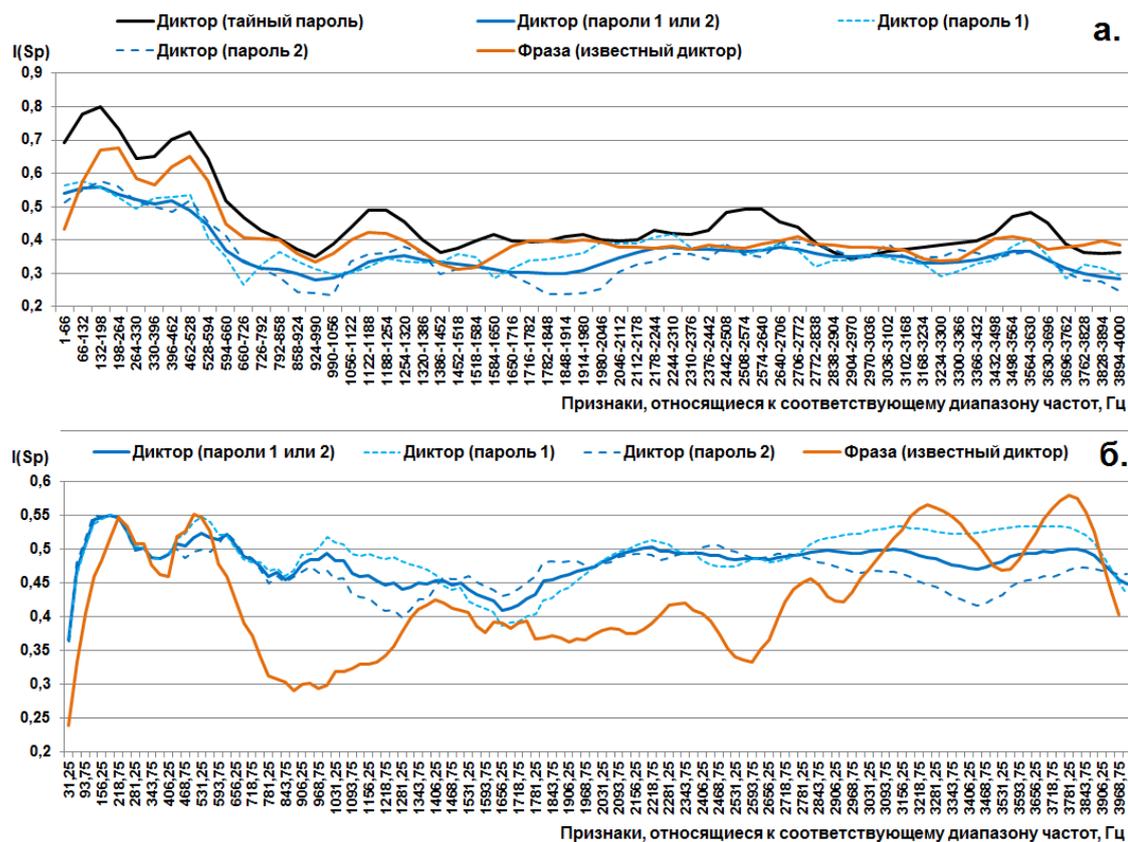


Рис. 1. Информативность признаков: а — первой группы, б — второй группы

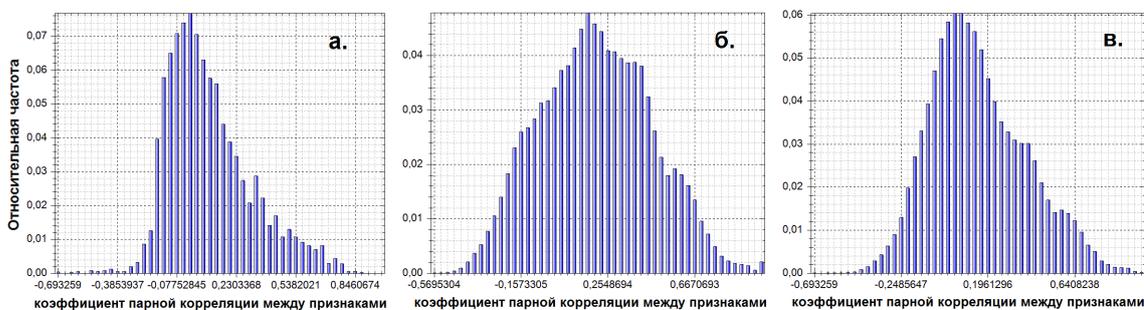


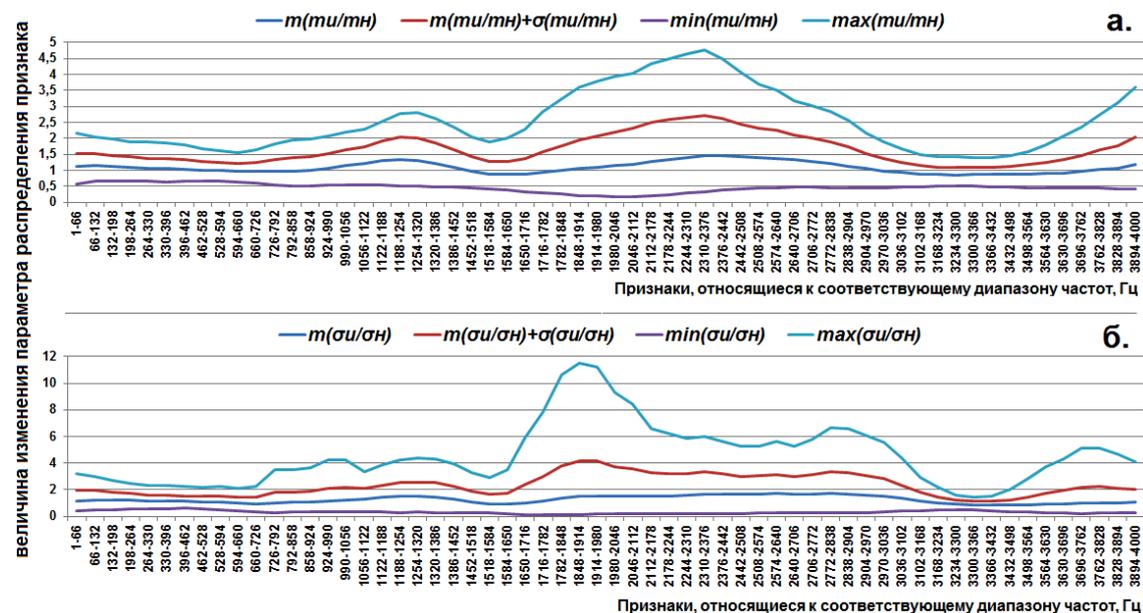
Рис. 2. Взаимная корреляционная зависимость: а — 100 признаков из первой группы, б — 512 признаков из второй группы, в — 100 признаков из первой и 512 признаков второй группы

Отметим что информативность признаков из группы 1 для задачи распознавания парольных фраз выше, чем для задачи распознавания диктора. Вторая группа признаков, напротив, более информативна для задачи распознавания диктора, нежели произносимых фраз. Для случая распознавания диктора по тайным голосовым паролям (когда произносимые разными дикторами фразы отличаются) информативность всех рассматриваемых признаков значительно выше, на рис. 1 а можно видеть синергетический эффект от «суммирования информативности», возникающий при распознавании не только диктора, но и фразы, которую он произносит.

Описанные признаки имеют распределения значений, достаточно близкие к логнормальному и реже к нормальному законам (что проверялось критерием Хи-квадрат).

## 5. ОЦЕНКА ВЛИЯНИЯ ПФС ДИКТОРА НА ПАРАМЕТРЫ ЕГО ГОЛОСА И ИХ ИНФОРМАТИВНОСТЬ

Изменения голосовых признаков из первой группы, вызванные сменой ПФС диктора, носят волнообразный характер и иллюстрируются на рис. 3.



**Рис. 3.** Трансформация параметров распределения признаков группы 1 при смене ПФС диктора с нормального (н) на любое измененное (и): а — математических ожиданий  $m()$ , б — среднеквадратичных отклонений  $\sigma()$  ( $max$  и  $min$  — функции максимума и минимума)

Почти всех испытуемых (порядка 90 %) по характеру этих изменений можно разделить на 3 условных категории (рис. 4 а): субъекты с ярко выраженным пиком изменений в окрестности частот 1700–2700 Гц и двумя слабо выраженными (700–1300 Гц и 3200–3700 Гц), субъекты с двумя выраженными пиками в интервалах частот 1000–1400 Гц и 2500–2900 Гц, субъекты с тремя пиками — слабым (750–1000 Гц), сильным (1650–2100 Гц) и пиком в области частот 3650–4000 Гц. Во всех измененных состояниях для большинства испытуемых и всех произнесенных ими паролей признаки  $\bar{a}_1$  меняются схожим образом (рис. 4 б). Данный результат можно назвать неожиданным.

Направленность изменений признаков группы 2 для различных испытуемых примерно одинакова и носит достаточно стойкий характер (рис. 5). Изменения, происходящие со значениями голосовых параметров  $\bar{a}_2$ , во всех рассматриваемых стадиях опьянения диктора почти равнозначны. Однако эти изменения заметно отличаются от тех, что происходят в сонном состоянии. В области высоких частот признаки второй группы почти не меняются в случае, когда диктор находится в сонном состоянии, но существенно меняются при алкогольном опьянении (даже незначительном). На регистрации данных изменений (по сравнению с эталоном трезвого диктора) может строиться пространство признаков, используемых для распознавания алкогольного опьянения.



Рис. 4. Пример изменения средних значений признаков группы 1: а — для 3-х дикторов в сонном состоянии, б — для одного диктора в зависимости от состояния

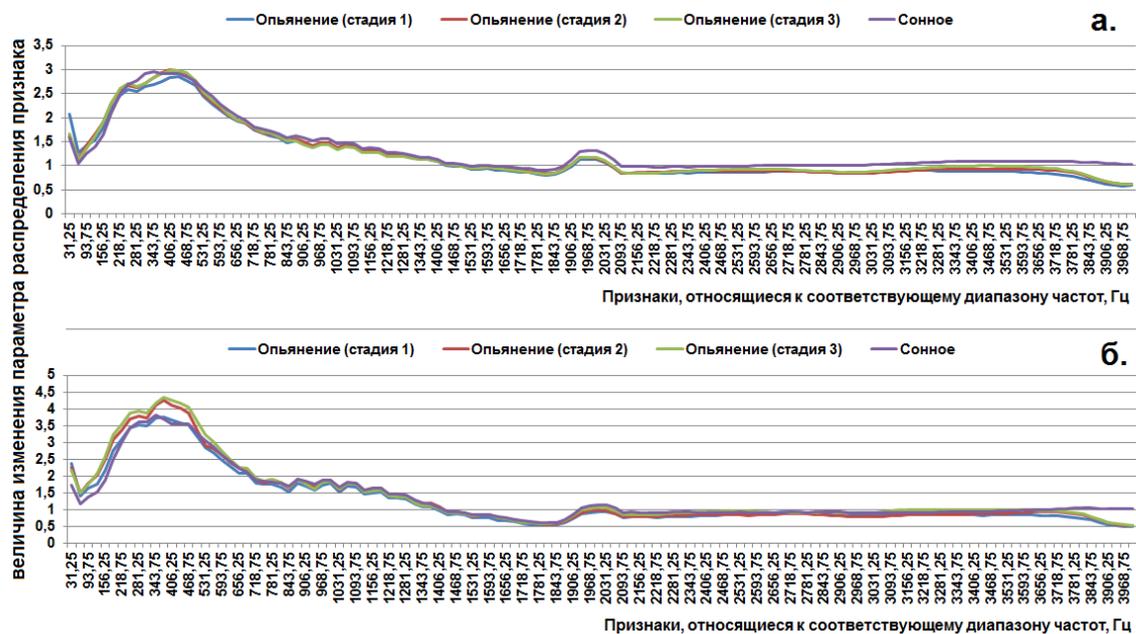


Рис. 5. Трансформация параметров распределения признаков группы 2 при смене ПФС диктора с нормального на измененное: а — математических ожиданий, б — среднеквадратичных отклонений

Описанные изменения голосовых параметров сказываются на их информативности (рис. 6). Почти все признаки первой группы, относящиеся к частотам свыше 600 Гц, становятся более информативными, если диктор находится в состоянии легкого опьянения (1 и 2 стадии). Этот результат также можно отнести к категории неожиданных. Формально приведенные данные о признаках  $\bar{a}_1$  указывают на то, что с их помощью можно было бы легче распознавать дикторов, если бы каждый из них всегда находился в состоянии легкого алкогольного опьянения, но, если ПФС диктора меняется, признаки также меняются очень значительно (рис. 3 и 4). Поэтому признаки группы 1, характеризующие частоту свыше 600 Гц, использовать для распознавания диктора не рекомендуется, так как они крайне неустойчивы к изменению ПФС своего владельца.

Информативность признаков второй группы, напротив, снижается после приема алкоголя. Прием естественных растительных средств седативного действия в целом ока-

зывает меньшее влияние на информативность биометрических параметров голоса, чем прием алкоголя.

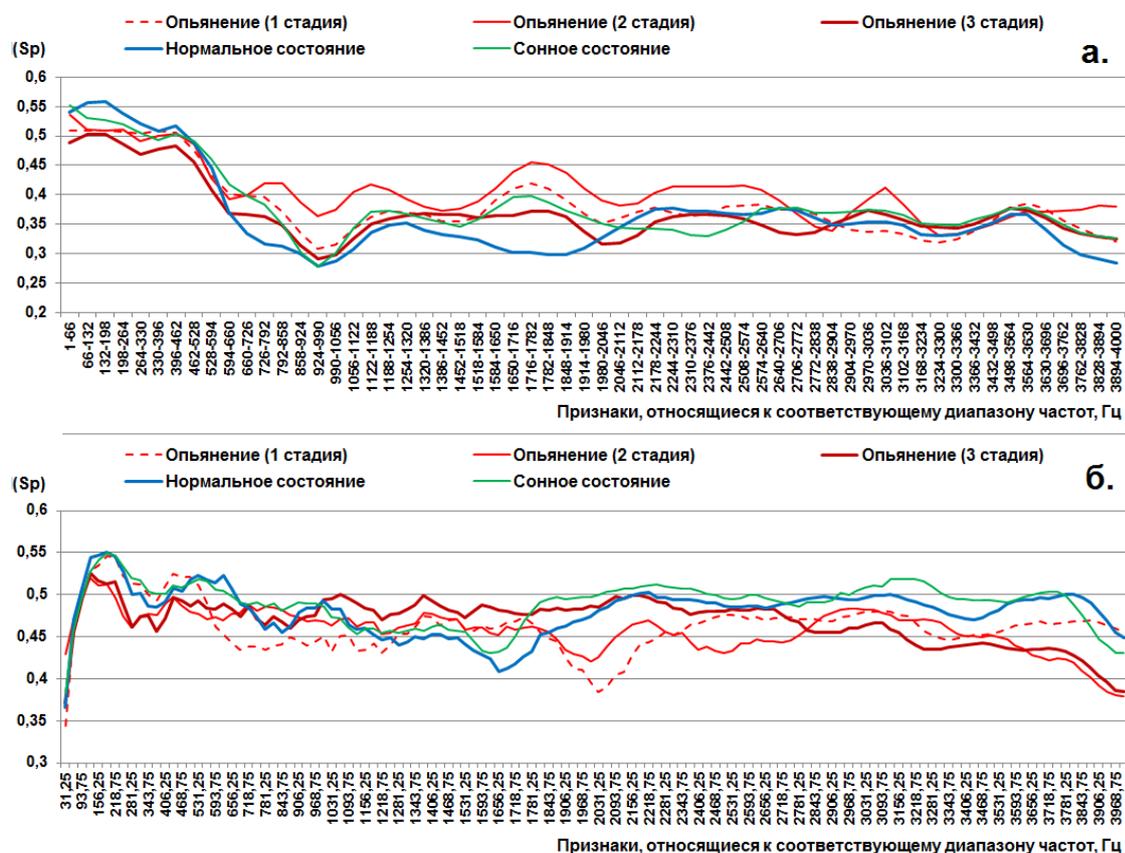


Рис. 6. Изменение информативности признаков в зависимости от ПФС диктора: а — первой группы, б — второй группы

## 6. РЕЗУЛЬТАТЫ РАСПОЗНАВАНИЯ ДИКТОРОВ ПО ГОЛОСУ

Надежность процедуры верификации биометрических образов определяется вероятностью (или процентом) ошибок 1-ого и 2-ого рода (False Rejection Rate (FRR) и False Acceptance Rate (FAR), соответственно). В настоящей работе критерием эффективности функционала (надежности принимаемых решений) считается средняя оценка вероятности ошибок 1-го и 2-го рода  $((FRR + FAR)/2)$ . Физический смысл критерия заключается в получении оценки средней вероятности ошибки любого рода при использовании того или иного решающего правила (меры близости). Вероятность ошибки 1-ого или 2-ого рода в серии проведенных опытов оценивалась как число ошибок определенного рода к числу соответствующих опытов (для FRR на вход подавались образцы голосовых паролей «Свой», для FAR — образцы «Чужой»). С учетом проведения прямого численного эксперимента выбранный критерий является более удобным, чем критерий минимума равновероятной ошибки ( $EER = FRR = FAR$ ). При используемых мерах близости (описанных далее) и подходе к подсчету FRR и FAR не всегда удастся найти точку строгого равенства  $FRR = FAR$ .

В биометрических приложениях верификации диктора и его состояний апробируются и находят применение различные подходы к распознаванию образов: нечеткие экстракторы [6], искусственные нейронные сети (ИНС) [8–12], статистические методы принятия решений (в частности, Байесовский классификатор [13] и др. [14]), вычисление расстояния (Евклида, Пирсона, Махаланобиса) в пространстве признаков от предъявленного образца до эталона [12, 15], метод  $k$ -ближайших соседей [15], векторное квантование [15], модель гауссовых смесей [15], машина опорных векторов [15]. В биометрических приложениях защиты информации рекомендуется выполнять требование ГОСТ Р 52633.0 не компрометировать биометрический эталон [9]. Среди перечисленных подходов только первые два позволяют выполнить это требование [9].

Метод «нечетких экстракторов» заведомо проигрывает многим другим подходом, в частности, нейронным сетям, так как в пространстве малоинформативных признаков (это как раз тот случай) дает высокий процент неверных решений [6]. Авторы настоящей статьи придерживаются идеологии построения так называемых «широких» нейронных сетей [16], базирующихся на нейронах, в основе которых лежат различные функционалы (меры близости). Сначала данные обрабатывает функционал нейрона, потом его значение отправляется в активационную функцию. В классическом персептроне функционалом является функция взвешенного суммирования. По сути, этот подход является вариантом обобщения ИНС, статистических методов и метрик для расчёта расстояний.

Быстрые не итерационные алгоритмы обучения «широких» ИНС являются рекордно устойчивыми и впервые предложены в России несколько лет назад для решения задач биометрической аутентификации [17]. Позже они легли в основу серии стандартов ГОСТ Р 52633. Данные алгоритмы не являются единственно возможными и порождают целый класс потенциально более эффективных алгоритмов. Направление развивается в основном силами ученых России и Казахстана.

Подход на базе «широких» сетей свободен от недостатков «глубоких» ИНС [16], связанных со сложностью обучения распознающего автомата и необходимостью использования большой обучающей выборки (сотни или тысячи примеров), и является более корректным применительно к настоящей задаче. Особенность «широкой» ИНС в том, что с увеличением количества нейронов не возрастает сложность обучения, что крайне важно. Каждый слой «широкой» ИНС строится исходя из того, что все нейроны могут ошибаться, но должны ошибаться по-разному. Таким образом, увеличение числа нейронов слоя не может повысить количества ошибок, а только линейно увеличивает время принятия решения. Однако если выходы нейронов сильно коррелированы, то наращивать их численность не имеет смысла.

Оставаясь в рамках концепции «широких» сетей, усовершенствование алгоритма принятия решений можно проводить почти бесконечно, дополняя конфигурацию ИНС нейронами на базе иных мер близости. Главное, чтобы каждый нейрон был уникален (использовал отличающийся функционал либо обрабатывал отличающиеся значения признаков). Чтобы в будущем сформировать высокоэффективную «широкую» сеть для данной задачи, требуется протестировать принципиально разные меры близости либо простые сети из них на устойчивость изменения состояния диктора.

В первую очередь решено использовать однослойные персептроны, обучаемые по ГОСТ Р 52633.5-2011 (процесс формирования нейросети и ее настройки подробно описан в стандарте и научных работах [9–11]), так как данный подход можно считать ориентиром для сравнения с другими.

Квадратичные формы (КФ) и их сети показывают хорошие результаты в ряде задач биометрической аутентификации [12] и апробировались ранее в задачах распознавания

ПФС по динамическим биометрическим признакам [13]. Мера Пирсона

$$\Pi = \sqrt{\sum_{j=1}^d \frac{(m_j - a_j)^2}{\sigma_j^2}}, \quad (8)$$

где  $d$  — размерность функционала ( $d \leq \eta$ ),  $m_j$  и  $\sigma_j$  — математическое ожидание и среднеквадратичное отклонение значений  $j$ -го признака, вычисляемые по данным обучающей выборки распознаваемого субъекта, является квадратичной формой, не учитывающей корреляцию между признаками. Аналогом этой меры, но позволяющим получить лучшие результаты в некоторых случаях [12] является мера хи-модуль:

$$\chi = \sum_{j=1}^d \frac{|m_j - a_j|}{\sigma_j^2}. \quad (9)$$

Также решено апробировать сети квадратичных форм (аналогов взвешенных мер Евклида)

$$\varepsilon = \sum_{j=1}^d (m_j - a_j)^2, \quad (10)$$

настраиваемых по адаптированному алгоритму ГОСТ Р 52633.5-2011, процесс обучения описан в работе [12].

Количество вычислительных элементов (нейронов) для персептронов и сетей квадратичных форм решено сделать равным 300 (дальнейшее увеличение количества нейронов не приводит к существенному улучшению качества решений и снижает производительность). Число входов (размерность  $d$ ) для всех нейронов менялось в процессе эксперимента (определялось оптимальное  $d$  по минимуму  $(FRR + FAR)/2$ ).

Следующие функционалы строятся на формуле гипотез Байеса. В литературе встречается 2 варианта ее использования в целях распознавания образов. Первый вариант — метод последовательного применения формулы Байеса (11) [13, 18] (МППФБ), который показал достаточно высокие результаты в ряде задач биометрической идентификации. Данный метод заключается в вычислении за  $d$  шагов апостериорных вероятностей гипотез, каждая из которых ассоциирована с определенным эталоном образа. В режиме верификации имеется 2 эталона: «Свой» и «Чужой», которые формируются по данным обучающей выборки. Первый строится по данным обучающей выборки распознаваемого субъекта, второй — на основе образцов других субъектов. На каждом  $j$ -м шаге за априорную вероятность принимается апостериорная вероятность, вычисленная на предыдущем шаге. На первом шаге гипотезы обычно считаются равновероятными, если нет статистических данных относительно них, в настоящей работе придерживались данного правила. Решение принимается в пользу гипотезы с наивысшей апостериорной вероятностью на последнем шаге:

$$P_h(a_{j-1}) = \frac{P_h(a_{j-1})p_h(a_j)}{\sum_{i=1}^{\Gamma} P_i(a_{j-1})p_i(a_j)}, \quad (11)$$

где  $P_h(a_j)$  — апостериорная вероятность  $h$ -й гипотезы, зависящая от  $j$ -го признака,  $p_h(a_j)$  — условная вероятность  $h$ -й гипотезы на  $j$ -м шаге принятия решений, равная плотности вероятности  $j$ -го признака,  $\Gamma$  — количество гипотез, в данном случае  $\Gamma = 2$ .

Второй вариант применения формулы Байеса заключается в конструировании многомерного функционала наибольшего правдоподобия (МФНПБ), который для случая верификации (2-х гипотез) можно представить в виде:

$$P_h(\bar{a}) = \frac{0,5 \prod_{j=1}^N p_h(a_j)}{\sum_{i=1}^{\Gamma} (0,5 \prod_{j=1}^N p_h(a_j))} = \frac{\prod_{j=1}^N p_h(a_j)}{\sum_{i=1}^{\Gamma} \prod_{j=1}^N p_h(a_j)}. \quad (12)$$

Решение в пользу  $h$ -й гипотезы также принимается по максимальной апостериорной вероятности  $P_h(\bar{a})$ .

Последние рассматриваемые меры близости основаны на критериях проверки закона распределения случайной величины Джини (13)–(14) и Крамера-фон Мизеса (15)–(16). На базе данных критериев в [8] предложены интегральные (13)–(15) и дифференциальные (14)–(16) функционалы для распознавания образов. Первые оперируют вероятностями, вторые — плотностями вероятностей значений признаков:

$$DgI = \int_{-\infty}^{+\infty} |P(\bar{a}) - \dot{P}(\hat{a})| d\hat{a}, \quad (13)$$

$$DgD = \int_{-\infty}^{+\infty} |p(\bar{a}) - \dot{p}(\hat{a})| d\hat{a}, \quad (14)$$

$$KfMI = \int_{-\infty}^{+\infty} (P(\bar{a}) - \dot{P}(\hat{a}))^2 d\hat{a}, \quad (15)$$

$$KfMD = \int_{-\infty}^{+\infty} (p(\bar{a}) - \dot{p}(\hat{a}))^2 d\hat{a}, \quad (16)$$

Мощность критерия хи-квадрат выше, чем мощность интегрального критерия Джини [8], но ниже мощности интегрального критерия Крамера-фон Мизеса [8], при этом доказано, что мощность дифференциального критерия Джини выше, чем интегрального [8, 15]. Для верифицируемого субъекта по данным обучающей выборки «Свой» вычисляются параметры распределения значений каждого признака, то есть признаки воспринимаются как набор случайных величин. Эти параметры представляют собой эталон. Далее исходные значения признаков нормируются по формуле

$$\bar{a} = \frac{a_j - m(a_j)}{\sigma(a_j)}, \quad (17)$$

после чего они могут быть представлены в виде значений одной нормально распределенной случайной величины с математическим ожиданием, равным нулю, и единичным среднеквадратичным отклонением. На этапе принятия решений значения признаков, вычисляемые из предъявленного биометрического образа, нормируются аналогичным образом (под параметры эталона образа, заявленного к верификации), и формируется эмпирическое распределение. Однако параметры этого распределения будут отличаться, поэтому при сравнении эмпирического и эталонного распределений с помощью любого из указанных критериев неизбежно будут регистрироваться отклонения. По величине этих отклонений предлагается принимать решение об аутентичности образа [8].

Функционалы (13), (14), (15) и (16) можно обучать не только на данных «Свой», но и на данных «Все чужие» (обучение «Свой-Чужой» по аналогии с МППФБ — каждому субъекту в этом случае соответствуют 2 эталона). Такой способ позволяет находить близость входного образа сразу к двум эталонам и принимать решение в пользу одной из двух гипотез «Свой» и «Чужой», сравнивая расчетные значения соответствующих функционалов. Это дает возможность не искать оптимальное пороговое значение мер близости (13), (14), (15) и (16). В случае дифференциальных функционалов данный подход дает меньшее число ошибок, в чем можно убедиться из таблицы 1.

Эталоны образов формировались по данным обучающих выборок, полученных в нормальном состоянии субъектов. Для формирования эталона «Свой» использовался 21 образец голоса одного диктора, при формировании эталона «Чужой» — 64 образца от различных дикторов (объем обучающей выборки задавался исходя из рекомендаций ГОСТ Р 52633.5-2011 и того факта, что увеличение объема отрицательно влияет на удобство использования системы распознавания голоса). Эксперимент проводился в 2-х вариантах: в первом случае на вход мерам близости подавались биометрические образы нормального состояния дикторов, во втором — измененных состояний. При проведении эксперимента в каждой сессии распознавания использовался определенный голосовой пароль (проведено всего 8 сессий), то есть предполагалось, что парольная фраза не является тайной. Результаты представлены в таблице 1.

**Таблица 1.** Наилучшие результаты по распознаванию субъектов и размерность, при которой они достигаются

Мера близости / название подхода	d / число входов нейрона	ПФС субъектов			
		нормальное		измененное	
		FRR	FAR	FRR	FAR
Мера Пирсона	45 ± 5 (7 ± 2)	0,04 (±0,005)	0,045 (±0,005)	0,269 (±0,01)	0,257 (±0,01)
Мера хи-модуль	45 ± 5 (7 ± 2)	0,042 (±0,005)	0,039 (±0,005)	0,246 (±0,01)	0,271 (±0,01)
Сеть КФ	5 ± 2 0,085	(±0,005)	0,055 (±0,005)	0,369 (±0,01)	0,357 (±0,01)
Персептрон	30 ± 10	0,053 (±0,005)	0,031 (±0,005)	0,29 (±0,01)	0,31 (±0,01)
DgI	120±20 (612)	0,046 (±0,005)	0,05 (±0,005)	0,331 (±0,01)	0,345 (±0,01)
DgD	612	0,141 (±0,01)	0,123 (±0,01)	0,408 (±0,01)	0,39 (±0,01)
KfMI	120 ± 20 (612)	0,243 (±0,01)	0,126 (±0,01)	0,31 (±0,01)	0,318 (±0,01)
KfMD	612	0,229 (±0,01)	0,195 (±0,01)	0,376 (±0,01)	0,364 (±0,01)
обучение «Свой-Чужой»					
МППФБ	612	0,051 (±0,005)	0,025 (±0,005)	0,709 (±0,01)	0,019 (±0,01)
МФНПБ	612	0,057 (±0,005)	0,023 (±0,005)	0,711 (±0,01)	0,018 (±0,01)
DgI	120 ± 20 (612)	0,018 (±0,005)	0,148 (±0,01)		
DgD	612	0,036 (±0,005)	0,102 (±0,01)		
KfMI	120 ± 20 (612)	0,028 (±0,005)	0,092 (±0,005)		
KfMD	612	0,08 (±0,005)	0,057 (±0,005)		

Эмпирически установлено, что при условии нахождения диктора в нормальном ПФС, оптимальным количеством признаков является  $\eta = 532$ , в том числе  $\eta_1 = 20$  (признаки с

частотой до 400 Гц) и  $\eta = 512$  (весь рассматриваемый частотный диапазон до 4000 Гц), по крайней мере, при использовании рассмотренных мер близости. Остальные признаки либо не вносят существенного вклада в процесс формирования решений описанными функционалами, либо способствуют накоплению ошибок и снижают надежность распознавания субъектов. Если диктор находится в измененном ПФС, то из второй группы рекомендуется использовать признаки, только характеризующие частоты 2100–3700 Гц, так как они наименее изменчивы (рис. 5).

Наилучшие результаты при распознавании диктора, находящегося в нормальном состоянии, показывают МППФБ и МФНПБ (количество ошибок менее 4 %). Данные функционалы дают высоко коррелирующие результаты, их особенностью является отсутствие насыщения (вероятность ошибочных решений всегда снижается при повышении размерности, если признак несет информацию). Другие меры близости ведут себя иначе, в частности мера Пирсона и хи-модуль (табл. 1). Тем не менее, все рассмотренные выше функционалы теряют мощность, если значения признаков существенно отклоняются от своего математического ожидания. При изменении ПФС диктора на этапе биометрической аутентификации (по сравнению с тем состоянием, которое было на этапе обучения системы) его распознавание в пространстве указанных признаков при помощи рассматриваемых мер близости становится затруднительным (количество ошибок превышает 25 %). Наиболее устойчивыми функционалами к изменению ПФС диктора являются мера Пирсона и хи-модуль.

Также проведен аналогичный эксперимент для случая тайных голосовых паролей. Субъекты делились на группы из 8 человек (по количеству паролей), среди которых проводились отдельные сессии по распознаванию дикторов (то есть создавалась база образов голоса). Распознавание проводилось только с помощью меры Пирсона и хи-модуль, наилучший результат показала мера хи-модуль:

- мера хи-модуль (нормальное ПФС дикторов): 0,023 ( $\pm 0,005$ ), 0,017 ( $\pm 0,005$ );
- мера хи-модуль (измененное ПФС дикторов): 0,171 ( $\pm 0,01$ ), 0,128 ( $\pm 0,01$ ).

Таким образом, число ошибок удалось снизить почти в 2 раза.

## 7. ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

По результатам исследований предлагаемые способы вычисления амплитудно-частотных характеристик голосовых паролей оказались очень чувствительны к изменению психофизиологического состояния диктора, по крайней мере, в классе таких воздействий, как прием алкоголя и естественных растительных средств седативного действия (пустырник, мята, валериана). При соблюдении условия нахождения лица, проходящего процедуру биометрической аутентификации, в адекватном состоянии (без предварительного употребления алкоголя) количество ошибок распознавания по фиксированной парольной фразе (из 2-х слов) составляет порядка 4 %, по тайной парольной фразе — 2 %. Если предварительно употребить алкоголь, число ошибочных решений повышается соответственно до 25–30 % и 15 %.

Повышение надежности распознавания дикторов видится возможным на пути объединения рассмотренных функционалов в сеть. Требуется совместно использовать только те функционалы, коррелированность решений которых невысока.

Установлено, что при нахождении субъекта в состоянии опьянения или сонном состоянии ряд признаков, описывающих частоты 1850–2100 Гц, существенно меняется. Признаки, описывающие частоты свыше 3700 Гц, меняются только в состоянии

опьянения. На регистрации этих отклонений может быть построен способ вычисления признаков, характеризующих состояния покоя, опьянения и сонное состояние.

### Список литературы

1. Богомолов А. В., Гридин Л. А., Кукушкин Ю. А., Ушаков И. Б. Диагностика состояния человека: математические подходы. М.: Медицина, 2003.
2. Ильин Е. П. Психофизиология состояний человека. СПб.: Питер, 2005.
3. De Waard D. The measurement of drivers' mental workload. Traffic Safety Research Centre VSC. University of Groningen, Haren, 1996.
4. Mascord, D. J. & Heath, R. A. Behavioral and physiological indices of fatigue in a visual tracking task // Journal of Safety Research. 1992. Vol. 23. P. 19–25.
5. Епифанцев Б. Н., Ложников П. С., Сулаво А. Е., Борисов Р. В. Комплексированная система идентификации личности по динамике подсознательных движений // Безопасность информационных технологий. 2011. № 4. С. 97–102.
6. Сулаво А. Е., Еременко А. В., Борисов Р. В. Генерация криптографических ключей на основе голосовых сообщений // Прикладная информатика. 2016. № 5, С. 76–89.
7. Еременко А. В., Сулаво А. Е., Мишин Д. В., Федотов А. А. Идентификационный потенциал клавиатурного почерка с учетом параметров вибрации и силы нажатия на клавиши // Прикладная информатика. 2017. № 1 (67). Т. 12. С. 79–94.
8. Иванов А. И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции. Пенза: Изд-во ПНИЭИ, 2016.
9. Ложников П. С., Сулаво А. Е., Еременко А. В., Волков Д. А. Экспериментальная оценка надежности верификации подписи сетями квадратичных форм, нечеткими экстракторами и перцептронами // Информационно-управляющие системы. 2016. № 5. С. 73–85.
10. Ахметов Б. С., Иванов А. И., Фунтиков В. А., Безяев А. В., Малыгина Е. А. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа. Монография. Алматы: ТОО «Издательство LEM», 2014.
11. Иванов А. И. Нейросетевая защита конфиденциальных биометрических образов гражданина и его личных криптографических ключей. Монография. Пенза, 2014
12. Сулаво А. Е., Еременко А. В., Толкачева Е. В., Борисов Р. В. Комплексирование независимых биометрических признаков при распознавании субъектов на основе сетей квадратичных форм, перцептронов и меры ХИ-модуль // Информационно-управляющие системы. 2017. № 1 (86). С. 50–62. doi:10.15217/issn1684-8853.2017.1.50.
13. Васильев В. И., Сулаво А. Е., Борисов Р. В., Жумажанова С. С. Распознавание психофизиологических состояний пользователей на основе скрытого мониторинга действий в компьютерных системах // Искусственный интеллект и принятие решений. 2017. № 3. С. 95–111.
14. Серикова Н. И., Иванов А. И., Серикова Ю. И. Оценка правдоподобия гипотезы о нормальном распределении по критерию Джини для числа степеней свободы, кратного числу опытов // Вопросы радиоэлектроники. 2015. № 1 (1). С. 85–94.
15. Первушин Е. А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование. 2011. Вып. 24. С. 41–54.
16. Сулаво А. Е. Сравнение функционалов на базе статистических критериев для формирования широких нейронных сетей сверхбыстрого обучения // Инфографика и информационный дизайн: визуализация данных в науке: Материалы Междунар. науч.-практ. конф. (Омск, 17–18 нояб. 2017 г.) / Минобрнауки России, ОмГТУ; SangMyungUniversity; [редкол.: О. В. Батенькина (отв. ред.) и др.]. Омск : Изд-во ОмГТУ, 2017.
17. Волчихин В. И., Иванов А. И., Фунтиков В. А. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации. Монография. Пенза: Изд-во Пензенского государственного университета, 2005.
18. Епифанцев Б. Н., Ложников П. С., Сулаво А. Е., Жумажанова С. С. Идентификационный потенциал рукописных паролей в процессе их воспроизведения // Автометрия. 2016. № 3. С. 28–36.

Поступила в редакцию 12.07.2017, окончательный вариант — 10.08.2017.

## INFLUENCE OF A SPEAKER'S PSYCHO-PHYSIOLOGICAL STATE TO HIS VOICE PARAMETERS AND RESULTS OF BIOMETRIC AUTHENTICATION BY SPEECH ENABLED PASSWORD

Sulavko A. E.<sup>1,2</sup>, Eremenko A. V.<sup>2</sup>, Borisov R. V.<sup>3</sup>, Inivatov D. P.<sup>1</sup>

<sup>1</sup>Omsk State Technical University, Omsk, Russia

<sup>2</sup>Candidate of Technical Sciences Omsk State Technical University, Omsk, Russia

<sup>3</sup>The Siberian Automobile and Highway University, Omsk, Russia

### Abstract

In this work, two methods were used to calculate the identification characteristics of the speaker's voice. One of them is based on the direct Fourier transform, the second — on the window transformation with the subsequent integration of the values of each harmonic of all the windows. The information content of these characteristics is determined. An estimation is given of how the parameters of the voice and their informativeness change depending on the degree of alcoholic intoxication of a person and in a sleepy state. A computational experiment was carried out to evaluate the reliability of recognition of speakers in the space of selected features using functionals based on the Bayesian hypothesis formula, Pearson measure, chi-module measure, Gini criterion, Cramervon Mises, and perceptrons trained in GOST R 52633.5-2011, and networks of quadratic forms. An estimation is given of the stability of these methods and functionals to the psychophysiological state of the speaker in terms of the robustness of the obtained recognition results.

**Keywords:** *pattern recognition, speech signal parameters, speech enabled password, biometric authentication, psychophysiological state of the speaker, state of alcoholic intoxication.*

**Citation:** A. E. Sulavko, A. V. Eremenko, R. V. Borisov & D. P. Inivatov, "Vliyaniye psikhofiziologicheskogo sostoyaniya diktora na parametry ego golosa i rezul'taty biometricheskoi autentifikatsii po rechevomu parolyu" [Influence of a Speaker's Psycho-physiological State to His Voice Parameters and Results of Biometric Authentication by Speech Enabled Password], *Computer tools in education*, no. 4, pp. 29–47, 2017 (in Russian).

*Received 12.07.2017, the final version — 10.08.2017.*

**Alexey E. Sulavko, Candidate of Technical Sciences Omsk State Technical University, assistant professor, [sulavich@mail.ru](mailto:sulavich@mail.ru)**

**Alexander V. Eremenko, Candidate of Technical Sciences Omsk State Transport University, design engineer, [nexus@mail.ru](mailto:nexus@mail.ru)**

**Roman V. Borisov, The Siberian Automobile and Highway University, lecturer, [brv1986@yandex.ru](mailto:brv1986@yandex.ru)**

**Daniil P. Inivatov, Omsk State Technical University, Radio Engineering Faculty, BIT-151, student, [daniilini@mail.ru](mailto:daniilini@mail.ru)**

**Сулавко Алексей Евгеньевич,**  
кандидат технических наук, доцент ОмГТУ,  
[sulavich@mail.ru](mailto:sulavich@mail.ru)

**Еременко Александр Валериевич,**  
кандидат технических наук,  
инженер-проектировщик ОмГУПС,  
[nexus-@mail.ru](mailto:nexus-@mail.ru)

**Борисов Роман Владимирович,**  
преподаватель СибАДИ,  
[brv1986@yandex.ru](mailto:brv1986@yandex.ru)

**Иниватов Даниил Павлович,**  
студент радиотехнического факультета  
(БИТ-151) ОмГТУ,  
[daniilini@mail.ru](mailto:daniilini@mail.ru)

© Наши авторы, 2017.  
Our authors, 2017.