

## Пилотный эксперимент по созданию базы русскоязычных речевых данных и распознаванию состояния утомления человека с использованием нейронной сети\*

© 2022 А.В. Яковлев<sup>1,2</sup>, В.О. Матыцин<sup>1,3</sup>, В.А. Велюга<sup>2</sup>, К.А. Найденова<sup>1</sup>,  
В.А. Пархоменко<sup>4</sup>

<sup>1</sup>Военно-медицинская академия

(194044, Санкт-Петербург, ул. Академика Лебедева, д. 6),

<sup>2</sup>Санкт-Петербургский государственный университет  
аэрокосмического приборостроения

(190000, Санкт-Петербург, ул. Большая Морская, д. 67)

<sup>3</sup>ФГБОУ ВО ПСПбГМУ им. И.П. Павлова Минздрава России  
(197022, Санкт-Петербург, ул. Льва Толстого, д. 6-8)

<sup>4</sup>Санкт-Петербургский политехнический университет Петра Великого  
(195251, Санкт-Петербург, ул. Политехническая, д. 29)

E-mail: sven-7@mail.ru

Поступила в редакцию: 15.11.2022

Качественные психофизиологические исследования сопряжены с созданием доступных и хорошо организованных баз данных, требующих большую предварительную работу по разработке измерительных комплексов, включающих не только средства для измерения психофизиологических параметров человека, но и его эмоционального состояния, которое отображается в выражении лица, речи и поведенческих паттернах респондентов. Измерительные комплексы должны также включать и средства обработки экспериментального материала. Суть исследования состояла в проведении пилотного эксперимента по созданию прототипа Базы Речевых Данных русскоязычных респондентов, получения ответов на методические вопросы, возникающие у специалистов при использовании базы для задачи распознавания состояния утомления человека. Разработан аппаратно-программный комплекс, позволяющий синхронно регистрировать психофизиологические параметры, видеозаписи поведенческих реакций и аудиозапись речи человека. В качестве модели физического утомления использовался кардиореспираторный тест с физической нагрузкой. До прохождения и после завершения теста добровольцы зачитывали набор стандартных фонетически представительных текстов. Полученные аудиозаписи обрабатывались с помощью специализированной нейронной сети, способной анализировать интегральные спектральные характеристики звука. Результаты пилотного эксперимента показали возможность распознавания состояния утомления человека по его речи, что позволяет перейти к созданию большого банка аудиозаписей и совершенствованию алгоритмов распознавания состояния утомления.

*Ключевые слова:* распознавание утомления, база речевых данных, инструментальный комплекс, кардио-респираторный тест, машинное обучение, нейронная сеть.

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Яковлев А.В., Матыцин В.О., Велюга В.А., Найденова К.А., Пархоменко В.А. Пилотный эксперимент по созданию базы русскоязычных речевых данных и распознаванию состояния утомления человека с использованием нейронной сети // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2022. Т. X, № Y. С. Z1–Z2. DOI: 10.14529/cmse220Y0Z.

---

\*Статья рекомендована к публикации программным комитетом научной конференции DAMDID 2022.

## Введение

Работа человека-оператора (далее – оператора) характеризуется напряжением внимания с необходимостью его переключения, а также нервно-психическим напряжением в связи с высокой ответственностью за результаты деятельности. Высокие нагрузки ведут к развитию у операторов состояния утомления, что сопряжено с угрозой пропуска значимых сигналов и немотивированного реагирования на сигналы ложные. Поэтому контроль за развитием утомления у специалистов операторского профиля в процессе их профессиональной деятельности является актуальной задачей. Однако решение данной задачи сопряжено с рядом проблем:

- моделирование условий труда специалиста в большинстве случаев не соответствует реальным условиям его деятельности;
- попытка провести какие-либо измерения в процессе трудовой деятельности с целью определить ее эффективность, представляет помеху работе оператора;
- прогностическая ценность существующих математико-статистических моделей оценки работоспособности оператора невелика в силу ограничений, накладываемых в большинстве случаев небольшими размерами изученных выборок.

В настоящее время активно развиваются методы распознавания состояния человека-оператора по речи, изображению и поведению, реализуемые с помощью систем автоматизированной оценки с применением нейронных сетей [1]. Такие системы позволяют оценивать состояния человека-оператора дистанционно, не отрывая его от привычной деятельности.

Для регистрации состояния утомления человека перспективным является речевой канал. Он прост и недорог и в наименьшей степени подвержен искажениям во время регистрации, по сравнению с записью видео или физиологических показателей.

Исследовательская активность, посвященная анализу речи человека, в основном сосредоточена в следующих направлениях:

- совершенствование алгоритмов обработки речевого сигнала, включая:
  1. совершенствование программных инструментов для визуализации речевого сигнала и расчета признаков, характеризующих речевой сигнал, в частности PRAAT [2, 3], ISIP [4], openSmile [5];
  2. адаптация современных алгоритмов машинного обучения для решения задач анализа речевого сигнала [1];
- разработка алгоритмов и методик распознавания различных состояний и патологий человека на основе анализа его речи, включая:
  1. распознавание различных эмоциональных состояний;
  2. распознавание ряда отклонений от нормальных психических и физиологических состояний;
  3. распознавание отдельных состояний человека, связанных с исполнением им своих функциональных обязанностей в процессе деятельности, в том числе состояния утомления [6].

Вместе с тем, основной проблемой для исследователей является наличие качественных наборов речевых данных или баз речевых данных (БРД), составляющих основу машинного обучения. Основная проблема состоит том, что для каждого языка необходимо создавать свои БРД. В частности, уже разработаны базы данных, содержащие большое количество записей речи дикторов, выражающих нейтральные, положительные либо отрицательные эмоции. Эти базы данных существуют в свободном доступе и служат в качестве источника

эталонных сигналов для распознавания эмоций в голосе. Однако эти БРД содержат английскую, немецкую, итальянскую речь [7, 8], при этом русскоязычной БРД с открытым доступом пока не представлено.

В настоящее время разработано значительное число алгоритмов обработки речи, изучены характеристики голоса, выявлены речевые параметры, которые способны варьировать в зависимости от функционального и эмоционального состояния человека. Однако создание алгоритмов, позволяющих распознать утомление человека по его речи в процессе профессиональной деятельности, находится на этапе разработки прежде всего по причине трудностей моделирования состояния утомления человека. Для создания такой БРД требуются ответы на следующие методические вопросы:

1. как моделировать утомление и что является достоверным критерием наступления состояния утомления;
2. какие тексты должны быть использованы для чтения;
3. какой длительности должны быть речевые фрагменты, достаточные для распознавания состояния утомления;
4. микрофоны какого качества необходимо использовать для записи речи и какой уровень «огрубления» исходных данных допустим при обучении.

Суть настоящей работы состояла в проведении пилотного эксперимента по созданию прототипа БРД русскоязычных респондентов с целью получения ответов на вышеперечисленные вопросы. Основные усилия были направлены на реализацию законченного процесса распознавания утомления, включающего подбор текстов для чтения респондентов, разработку аппаратно-программного инструментария, проведение самого эксперимента, организацию регистрируемых данных, формирование обучающей выборки с речевыми сигналами респондентов, ее преобразование в прототип БРД и, в меньшей степени, на анализ и выбор алгоритмов распознавания речи, так как в этом вопросе возможно опираться на уже готовые и проверенные модели.

В разделе 1 рассматриваются основные методы и средства извлечения первичной информации. Далее в разделе 2 приведена краткая характеристика разработанного комплекса. Разделы 3–5 посвящены дизайну, обработке и обсуждению результатов проведения пилотного эксперимента соответственно. Основные выводы работы изложены в разделе Заключение.

## **1. Материалы и методы**

Основу эксперимента составила синхронная регистрация речи в процессе чтения добровольцем стандартных фонетически представительных текстов. В пилотном эксперименте приняло участие 9 добровольцев (здоровые мужчины в возрасте 22–25 лет), подписавших информированное согласие. Один из добровольцев принял участие в эксперименте дважды. В ходе каждого исследования доброволец читал три стандартных текста («командный текст», «проза», «стих») до и после нагрузки. Таким образом, было получено 30 исходных аудиозаписей.

Для моделирования утомления использовали кардиореспираторный тест (КаРен) с максимальной физической нагрузкой, выполняемый добровольцами на велоэргометре Ergoline, при этом контролировали кардиореспираторные и метаболические параметры добровольцев с помощью эргоспирометрического комплекса MetaLyser (Cortex, Германия).

Для записи речи одновременно использовалось два микрофона: профессиональный миниатюрный петличный микрофон AKG C 417<sup>III</sup> («левый» канал) и высокочувствительный метрологический микрофон РСВ 378А14 (№ 61720-15 в Государственном реестре средств измерений) совместно с усилителем РСВ 482С («правый» канал). Для аналого-цифрового преобразования звукового сигнала микрофонов использовалась внешняя двухканальная звуковая карта M-AUDIO M-Track Plus (МКII).

Для распознавания рассматриваемых состояний утомления добровольцев использовалась глубокая нейронная сеть с топологией автоэнкодера реализованная в библиотеке auDeep [9]. Автоэнкодер реализован с помощью библиотеки TensorFlow версии 1.15. Обучение глубокой нейронной сети выполнялось на графической карте NVIDIA Quadro M4000.

Помимо перечисленного оборудования, в исследовании были задействованы полиграф Диана-7М и система видеозаписи Диана-Видео (Поликониус, Москва), с помощью которых регистрировали физиологические показатели, поведенческие и мимические реакции добровольцев. Также с помощью оригинального программного обеспечения регистрировали простую зрительно-моторную реакцию (ПЗМР) у добровольцев до и после предъявления им КаРен-теста. Данные исследования были включены в сценарии эксперимента, а собранные с их помощью показатели поступали в созданную базу данных для хранения с целью их последующей обработки вне рамок настоящего пилотного эксперимента.

## **2. Краткая характеристика разработанного комплекса для формирования БРД**

Комплекс построен по архитектуре «клиент-сервер» и состоит из нескольких элементов, объединенных в локальную компьютерную сеть, включающую базу данных (модуль data-collection-system), размещенную на отдельном сервере в СУБД MySQL 5.8 и содержащую как метаданные эксперимента, так и все данные о респондентах, включая их речевые данные.

Для доступа к таблицам и полям базы данных был разработан web-сервис:

- терминал оператора для удалённого запуска процедур записи (модуль operator-cli);
- WebSocket-сервер, осуществляющий запись аудио по командам оператора (модуль audio-recorder);
- WebSocket-сервера, выполняющий показ слайдов с текстом для чтения на проекторе (модуль shower-server);
- Web-сервис, осуществляющий выгрузку обучающей выборки данных из БД (модуль data-collection-system) по определенным условиям: типы файлов, характеристики респондентов (выборка по полу, возрасту), состояния утомления респондентов, используемые тексты для чтения.

Сформированная база данных содержит: описания классов состояний утомления; тексты, читаемые добровольцами; описания сценариев выполнения экспериментов; данные о добровольцах; описание каждого эксперимента: реализуемый сценарий, идентификатор добровольца, дата проведения.

Для доступа к таблицам и полям базы данных разработан web-сервис, структура которого представлена на рис. 1. Работа с ним производится в следующей последовательности. На первом этапе организатор исследования указывает классы состояний утомления, подлежащие исследованию в рамках создаваемого сценария (дизайна) исследования. Далее, на втором этапе, в базу данных заносится текст, который будут читать добровольцы до и

после предъявления им нагрузки. На третьем этапе производится описание сценария (дизайна) исследования в виде матрицы, в которой построчно указывается порядок (алгоритм) регистрации показателей добровольца и предъявляемая ему нагрузка.

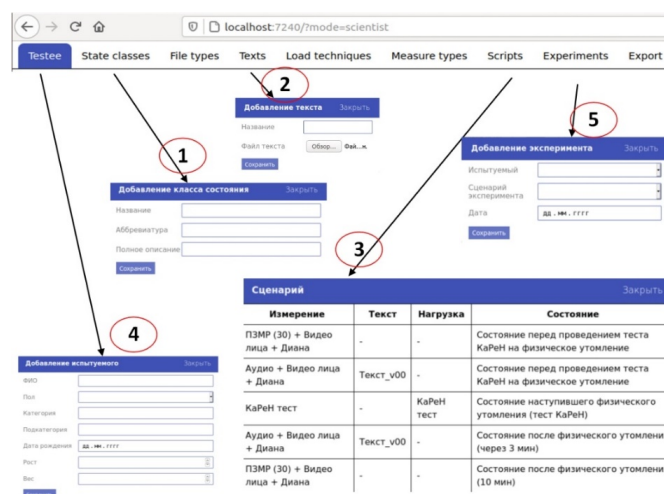


Рис. 1. Структура web-сервиса для доступа к базе данных

В представленном примере первая строка показывает, что комплекс должен провести у добровольца тест «простая зрительно-моторная реакция» (ПЗМР) с одновременной регистрацией видеоизображения лица (Видео лица) и физиологических показателей (Диана). При этом испытуемый находится в состоянии «не утомлен». Вторая строка показывает, что комплекс должен предъявить добровольцу для прочтения текст «Текст\_v00» и одновременно провести регистрацию его речи (Аудио), видеоизображения лица (Видео лица) и физиологических показателей (Диана). При этом испытуемый также находится в состоянии «не утомлен».

Третья строка показывает, что добровольцу предъявляется нагрузка в виде теста КаРен на физическое утомление. После этого, комплекс выполняет строки, аналогичные первым двум, но уже фиксирует состояние «утомлен». На четвертом этапе в комплекс заносятся данные добровольцев, принимающих участие в исследовании. На пятом этапе в соответствии с программой исследования, каждому добровольцу назначается дата эксперимента и «подцепляется» используемый сценарий эксперимента.

### 3. Дизайн пилотного эксперимента

В качестве моделей утомления обычно используются модель депривации сна [10], а также модель предъявления ментальной нагрузки, например тесты на переключение внимания либо решение арифметических задач в течение 100–180 минут [11–13]. Таким образом, для моделирования умственного утомления требуется значительное время, кроме того, методы моделирования умственной нагрузки сложно стандартизировать. Поэтому в данной работе было использовано физическое утомление, моделировать которое достаточно просто, проведя тест с максимальной физической нагрузкой «до отказа».

Данная модель не требует затрат большого количества времени. Было показано, что электромиографические признаки утомления мышц при физической нагрузке соответствуют достижению человеком анаэробного порога [14]. Любая нагрузка, умственная либо физическая, вызывает явления утомления, характеризующиеся изменением функционального

состояния организма человека с преобладанием процессов возбуждения либо (чаще всего) торможения в центральной нервной системе. Эти процессы оказывают влияние на свойства речи человека, поскольку речь регулируется центральной нервной системой.

Таким образом, модель физического утомления на пилотном этапе может считаться адекватной, при этом наиболее просто воспроизводимой. В качестве нагрузочного устройства в кардиореспираторном тесте (КаРен) использовали велоэргометр, поскольку он позволяет наиболее точно дозировать физическую нагрузку. Тест включал в себя фазы покоя (2 минуты), разогрева (2 минуты), нагрузки (индивидуальное время) и восстановления (1 минута). Доброволец крутил педали со скоростью 60 об/мин. Исходная нагрузка составила 20 Вт, затем она плавно возрастала со скоростью 20 Вт/мин. Контролировали достижение анаэробного порога (АП), критериями которого считали: 1) вентиляция по углекислому газу  $VCO_2$  начинает превышать вентиляцию по кислороду  $VO_2$ , соответственно значение дыхательного коэффициента (ДК) становится более 1; 2) вентиляционный эквивалент по кислороду  $VE/VO_2$ , бывший относительно постоянным, начинал расти за счет гипервентиляции; достижение точки перекреста линий регрессии  $VO_2$  и  $VCO_2$  за счет роста  $VCO_2$  [15]. Доброволец выполнял работу на велоэргометре до достижения им максимально переносимой физической нагрузки, при этом достижение анаэробного порога добровольцем расценивалось как объективное подтверждение развития у него состояния физического утомления.

Все исследования проводились в утренние часы, в специально оборудованном помещении. Все посторонние шумы в это время были устранены. Добровольцы приходили отдохнувшими. Перед началом каждого исследования проводилось их анкетирование с помощью анкеты САН (самочувствие, активность, настроение). В случае плохого самочувствия доброволец к исследованию не допускался.

В ходе пилотного эксперимента регистрировали два состояния добровольца: состояние перед моделированием физического утомления (состояние «не утомлен», далее – «НУ», Sf\_1) и состояние после физического утомления (через 3 минуты) (состояние «утомлен», далее – «У», Sf\_3). Перед предъявлением нагрузочного теста и через 3 минуты после его завершения доброволец читал специально подготовленный текст.

Выбор текста для чтения добровольцем представлял отдельную исследовательскую задачу. На момент исследования не было каких-либо достоверных сведений о том, чтение какого типа текста (командного, стихотворного и т.д.) может быть чувствительно к выявлению состояния «утомлен». Поэтому был сформирован единый текст, состоящий из небольшой первой «прогревочной» части и трех целевых частей:

- часть 1 («прогревочный текст») — было предъявлено несколько команд из ГОСТ 16600-72 для оценки средств связи (размер — 131 знак);
- часть 2 («командный текст») — было предъявлено большее количество команд из того же ГОСТ для оценки средств связи;
- часть 3 («проза») — был предъявлен фрагмент фонетически представительного текста;
- часть 4 («стих») — был предъявлен фрагмент фонетически представительного стихотворного текста.

Исходный текст автоматически разбивался на небольшие и хорошо видимые добровольцу фрагменты, которые тот читал. По мере чтения оператор комплекса давал команду на предъявление следующего фрагмента таким образом, чтобы не снижался темп чтения. При смене каждого фрагмента автоматически записывалось время относительно начала аудио-

записи. Предложенный подход хорошо себя зарекомендовал для случая чтения и дальнейшего разделения текстов разных типов внутри единого текстового документа, предъявляемого для чтения.

## 4. Обработка результатов пилотного исследования

По завершении экспериментальных исследований и проверки корректности собранных экспериментальных данных была проведена выгрузка требуемого набора данных из модуля data-collection-system с помощью web-сервиса. Были выгружены непосредственно файлы аудиозаписей и метки, относящие эти аудиофайлы к состояниям «утомлен» и «не утомлен».

Для обучения нейронной сети полученные аудиозаписи речи каждого из добровольцев как для состояния «НУ», так и для состояния «У», были разделены на 4 аудиофрагмента:

- аудиофрагмент № 1 – командный текст (далее — «Команды»): для обучения нейронной сети используется только записанный командный текст (прочтенная добровольцем и записанная часть 2 исходного текста);
- аудиофрагмент № 2 – фонетически представительная проза (далее — «Проза»): для обучения нейронной сети используется только записанная фонетически представительная проза (прочтенная добровольцем и записанная часть 3 исходного текста);
- аудиофрагмент № 3 – стихотворный текст (далее — «Стихи»): для обучения нейронной сети используется только записанный стихотворный текст (прочтенная добровольцем и записанная часть 4 исходного текста);
- аудиофрагмент № 4 – смешанный текст (далее — «Смешанный»): для обучения нейронной сети используются выбранные экспериментатором секунды из всех четырёх видов текстов, включая «прогревочный».

Результат «выгрузки» данных из БРД представлен на рис. 2.

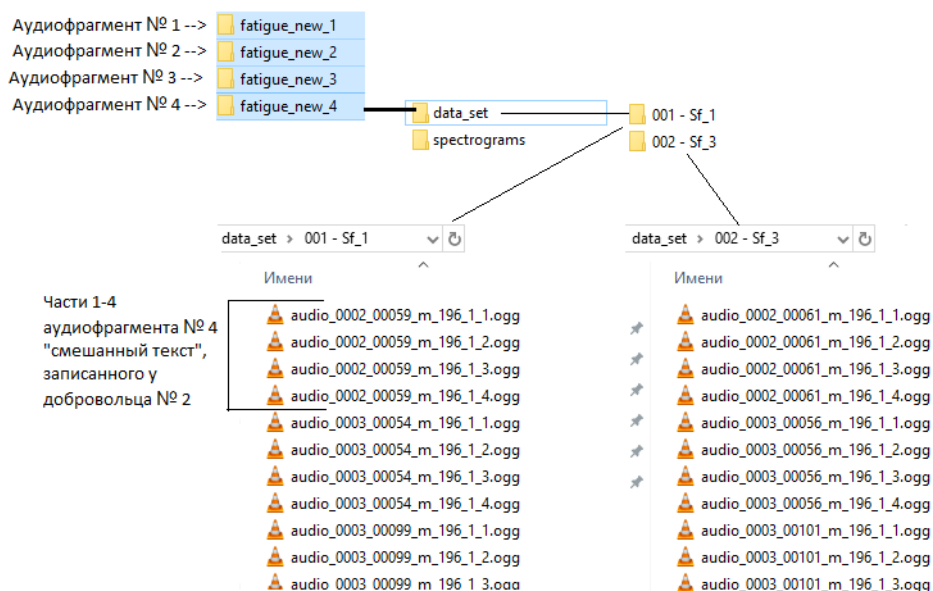


Рис. 2. Структура выгружаемых комплексом данных

Каждому из рассматриваемых аудиофрагментов соответствует свой каталог. В свою очередь, внутри каждого каталога размещаются два подкаталога «data\_set», содержащие интересующую выборку для обучения, и «spectrograms», куда в дальнейшем будут помещены сформированные библиотекой auDeer спектрограммы, используемые для обучения автоэн-

кодера. Подкаталог «data\_set» содержит два подкаталога, соответствующие двум классам оцениваемого состояния утомления: подкаталог «001 – Sf\_1» содержит аудиофайлы в формате .ogg, соответствующие состоянию добровольца «не утомлен», подкаталог «002 – Sf\_3» содержит аудиофайлы в формате .ogg, соответствующие состоянию добровольца «утомлен». В связи с тем, что тексты предъявлялись добровольцам фрагментами, то каждая из аудиозаписей также разбита на такие же фрагменты, средней длительностью более 11 секунд. Таким образом, для обучения нейронной сети было предъявлено по 40 аудиозаписей для каждого из состояний утомления для каждого аудиофрагмента.

Для решения задачи распознавания утомления человека была использована глубокая нейронная сеть с топологией автоэнкодера, реализованная в библиотеке auDeer [16]. Выбор указанной библиотеки определялся несколькими обстоятельствами. Во-первых, она показала высокую точность классификации акустических сцен конкурса IEEE AASP по обнаружению и классификации акустических сцен и событий (DCASE 2017) [9]. Во-вторых, она осуществляет значительный объём преобразований, связанных с корректным преобразованием исходных аудиофайлов в изображения, поступающие на вход библиотеки TensorFlow 1.15 и, соответственно, с экспортом сгенерированных признаков в формат CSV/ARFF. Использование библиотеки auDeer происходит в пять этапов (рис. 3):

- извлечение спектрограммы (англ. preprocess): извлечение спектрограмм и метаданных из необработанных аудиофайлов;
- обучение автоэнкодера (англ. train): обучение DNN на извлечённых спектрограммах;
- генерация признаков (англ. generate): генерация признаков с использованием обученного DNN;
- оценка признаков (англ. evaluate): оценка сгенерированных признаков;
- экспорт (англ. export): экспорт сгенерированных признаков в формат CSV/ARFF.



**Рис. 3.** Этапы использования библиотеки auDeer [1]

При обучении глубокой нейронной сети в библиотеке auDeer были установлены следующие метапараметры: window-width=0.08; window-overlap=0.04; mel-spectrum=128; epoch\_count=64; batch\_size=256; learning-rate=0.001; keep-prob=0.8; layer\_count=2; unit\_count=256. Варьировались следующие параметры обучения глубокой нейронной сети:

- Параметр № 1: канал регистрации: «левый» (качественный петличный микрофон), «правый» (высокочувствительный ненаправленный метеорологический микрофон).
- Параметр № 2: вид текста, аудиозапись которого используется для обучения: «команда», «проза», «стихи», смешанный.
- Параметр № 3: длительность используемой для обучения аудиозаписи отрывка текста: 7, 8, 9, 10, 11 сек.
- Параметр № 4: обрезка амплитуд записанных аудиоданных ниже указанного значения в децибелах (т. е. уровень «огрубления» исходных данных): clip\_below = -45 dB; clip\_below = -60 dB; clip\_below = -75 dB.



Результаты оценки точности обучения нейронной сети - матрицы смежности (англ. confusion matrix) для clip\_below=-45dB, clip\_below=-60dB, clip\_below=-75dB приведены, соответственно, в таблицах для петличного микрофона (табл. 1–3) и для ненаправленного метрологического микрофона (табл. 4–6). В каждой матрице отражены вероятности различия между состояниями «НУ» и «У», при этом «bold» шрифтом отмечены наилучшие значения правильных определений данных состояний, курсивом – умеренные значения правильных определений. В таблицах также приведено время обучения для каждого сочетания параметров (Т обучения).

**Таблица 1.** Матрицы смежности для clip\_below=-45dB с петличного микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.69	0.3	0.7	0.3	0.59	0.43	0.71	0.29
	У	0.33	0.67	0.51	0.49	0.33	0.67	0.41	0.59
	Время обучения	2m53s		2m56s		2m58s		2m56s	
8	НУ			0.71	0.29	0.59	0.4	0.82	0.18
	У			0.32	0.69	0.41	0.58	0.3	0.7
	Время обучения	3m10s		3m14s		3m15s			
9	НУ			<b>0.68</b>	<b>0.32</b>	0.68	0.32	0.81	0.2
	У			<b>0.27</b>	<b>0.73</b>	0.38	0.62	0.26	0.74
	Время обучения	3m32s		3m35s		3m33s			
10	НУ			0.68	0.33	0.65	0.35	0.73	0.27
	У			0.27	0.73	0.26	0.74	0.36	0.64
	Время обучения	3m51s		3m56s		3m53s			
11	НУ					0.6	0.4	0.62	0.38
	У					0.36	0.64	0.41	58
	Время обучения					4m20s		4m13s	

## 5. Обсуждение результатов

Исходя из табл. 1–6, точность распознаваний У и НУ достигала 69–80%. Несмотря на невысокую точность распознавания, результаты проведения обучения нейронной сети позволили сделать следующие выводы:

- При использовании аудиоданных с петличного микрофона («левый» канал) наилучший результат определения состояния был при использовании девятисекундного фрагмента прозы с параметром clip\_below равному -60dB. Нейросеть определяла верно состояние «не утомлен» в 69% случаев, а состояние «утомлен» в 79% случаев.
- При использовании аудиоданных с метрологического микрофона («правый» канал) наилучший результат определения состояния был при использовании также девятисекундного фрагмента прозы, но с параметром clip\_below равному -45dB. Нейросеть определяла верно состояние «не утомлен» в 80% случаев, а состояние «утомлен» в 75% случаев.

Таким образом, лучшие результаты приходятся на аудиофрагменты 8-9 секунд. При меньшей длительности, вероятно, нейросети не хватает данных для того, чтобы найти признаки для различия состояний, а при использовании фрагментов с большей длительностью

**Таблица 2.** Матрицы смежности для clip\_below=-60dB с петличного микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.6	0.4	0.61	0.39	0.55	0.46	0.55	0.46
	У	0.35	0.64	0.34	0.6	0.37	0.64	0.35	0.65
	Время обучения			2m53s		3m12s			
8	НУ			0.61	0.39	0.55	0.45	0.68	0.32
	У			0.31	0.69	0.47	0.53	0.36	0.64
	Время обучения			3m17s		3m15s			
9	НУ			<b>0.69</b>	<b>0.3</b>	0.54	0.46	0.7	0.29
	У			<b>0.21</b>	<b>0.79</b>	0.41	0.59	0.33	0.67
	Время обучения			3m39s		3m28s			
10	НУ			0.59	0.42	0.58	0.41	0.57	0.43
	У			0.36	0.65	0.42	0.57	0.47	0.53
	Время обучения					3m55s			
11	НУ					0.47	0.54	0.67	0.33
	У					0.43	0.57	0.46	0.55
	Время обучения								

**Таблица 3.** Матрицы смежности для clip\_below=-75dB с петличного микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.64	0.36	0.61	0.39	0.52	0.48	0.52	0.48
	У	0.37	0.63	0.4	0.6	0.36	0.64	0.32	0.69
	Время обучения			2m56s		2m57s			
8	НУ			0.71	0.28	0.62	0.37	0.68	0.33
	У			0.4	0.6	0.35	0.65	0.39	0.61
	Время обучения			3m13s		3m18s			
9	НУ			<b>0.66</b>	<b>0.34</b>	0.55	0.45	0.57	0.42
	У			<b>0.32</b>	<b>0.69</b>	0.42	0.57	0.36	0.63
	Время обучения					3m52s			
10	НУ			0.59	0.42	0.58	0.41	0.57	0.43
	У			0.36	0.65	0.42	0.57	0.47	0.53
	Время обучения					3m55s			
11	НУ					0.53	0.47	0.71	0.28
	У					0.46	0.54	0.52	0.48
	Время обучения								

происходит, по всей видимости, «переобучение» нейронной сети – явление, когда построенная модель хорошо объясняет примеры из обучающей выборки, но относительно плохо работает на примерах, не участвовавших в обучении. Таким образом, была показана способность разработанной модели распознавать в целом с вероятностью более 60% развитие физического утомления у добровольцев в условиях предъявления им физической нагрузки.

**Таблица 4.** Матрицы смежности для clip\_below=-45dB с ненаправленного метрологического микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.66	0.34	0.78	0.23	0.73	0.27	0.64	0.36
	У	0.33	0.67	0.4	0.6	0.37	0.63	0.34	0.66
	Время обучения	2m49s		2m48s		2m57s		3m5s	
8	НУ			0.8	0.2	0.61	0.39	0.57	0.42
	У			0.32	0.69	0.46	0.54	0.32	0.68
	Время обучения	3m8s		3m17s		3m21s			
9	НУ			<b>0.8</b>	<b>0.2</b>	0.64	0.36	0.73	0.27
	У			<b>0.25</b>	<b>0.75</b>	0.32	0.68	0.29	0.71
	Время обучения	3m33s		3m41s		3m36s			
10	НУ			0.84	0.16	0.68	0.33	0.66	0.34
	У			0.46	0.54	0.3	0.69	0.36	0.64
	Время обучения	4m0s		3m59s		4m7s			
11	НУ					0.66	0.34	0.71	0.28
	У					0.39	0.61	0.55	0.46
	Время обучения	4m27s		4m19s					

**Таблица 5.** Матрицы смежности для clip\_below=-60dB с ненаправленного метрологического микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.6	0.4	0.7	0.29	0.66	0.34	0.57	0.43
	У	0.32	0.69	0.41	0.58	0.46	0.55	0.34	0.66
	Время обучения	2m50s		2m57s		3m8s		2m58s	
8	НУ			0.84	0.16	0.6	0.39	0.83	0.17
	У			0.41	0.59	0.42	0.57	0.39	0.61
	Время обучения	3m9s		3m14s		3m33s			
9	НУ			<b>0.79</b>	<b>0.21</b>	0.55	0.46	0.59	0.41
	У			<b>0.28</b>	<b>0.72</b>	0.43	0.56	0.32	0.68
	Время обучения	3m32s		3m40s		3m38s			
10	НУ			0.73	0.27	0.56	0.44	0.53	0.47
	У			0.32	0.69	0.33	0.67	0.33	0.67
	Время обучения	4m2s		3m57s		4m1s			
11	НУ					0.59	0.41	0.66	0.34
	У					0.41	0.59	0.62	0.37
	Время обучения	4m26s		4m27s					

Дальнейшее совершенствование разрабатываемой методики оценки состояния утомления по речи будет продолжено в следующих направлениях:

- формирование полноценной БРД для русскоязычных респондентов;
- добавление для распознавания состояния утомления видеосигналов;

**Таблица 6.** Матрицы смежности для clip\_below=-75dB с ненаправленного метрологического микрофона

Длина текста, с	Тип текста →	Команды		Проза		Стихи		Смешанный	
	Состояния	НУ	У	НУ	У	НУ	У	НУ	У
7	НУ	0.68	0.32	0.68	0.33	0.66	0.34	0.62	0.37
	У	0.45	0.56	0.34	0.67	0.43	0.56	0.41	0.59
	Время обучения	2m59s		3m8s		3m6s		3m0s	
8	НУ			<b>0.78</b>	<b>0.22</b>	0.56	0.44	0.8	0.2
	У			<b>0.3</b>	<b>0.69</b>	0.45	0.55	0.42	0.57
	Время обучения			3m8s		3m12s		3m32s	
9	НУ			0.69	0.32	0.48	0.52	0.61	0.39
	У			0.34	0.67	0.38	0.62	0.36	0.64
	Время обучения			3m37s		3m38s		3m38s	
10	НУ			0.71	0.29	0.52	0.48	0.58	0.41
	У			0.32	0.69	0.33	0.67	0.38	0.62
	Время обучения			4m14s		4m8s		4m7s	
11	НУ					0.65	0.35	0.65	0.35
	У					0.47	0.53	0.56	0.43
	Время обучения					4m15s		4m26s	

- совершенствование алгоритма для распознавания состояния утомления;
- разработка методики оценки точности распознавания утомления.

## Заключение

Был проведен пилотный эксперимент с целью получения ответов на ряд методических вопросов, возникающих при создании БРД по утомлению. Для создания прототипа БРД был специально разработан аппаратно-программный комплекс для проведения исследований по распознаванию умственного и физического утомления. Была разработана методика формирования у добровольцев состояния физического утомления. Проведенный эксперимент позволил получить ответы на поставленные в начале исследования методические вопросы:

1. для моделирования физического утомления целесообразно использовать кардиореспираторный тест с максимальной физической нагрузкой, достоверным критерием наступления состояния утомления служит достижение добровольцем анаэробного порога при выполнении этого теста;
2. при формировании БРД для чтения целесообразно использовать тексты, соответствующие особенностям речевой коммуникации операторов, т.е. командные в сочетании с текстами типа «проза».
3. регистрируемые речевые фрагменты, достаточные для распознавания состояния утомления, должны быть не менее 8-9 секунд;
4. при использовании петличного микрофона целесообразно использовать «огрубление» -60dB, а при использовании более чувствительного микрофона -45dB.

## Литература

1. Zhang X.-J., Gu J.-H., Tao Z. Research of detecting fatigue from speech by PNN // Vol. 2. 2010. P. V2278–V2281. DOI: 10.1109/ICINA.2010.5636509.
2. Krajewski J., Batliner A., Golz M. Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach // Behavior Research Methods. 2009. Vol. 41, no. 3. P. 795–804. DOI: 10.3758/BRM.41.3.795.
3. Krajewski J., Trutschel U., Golz M., *et al.* Estimating fatigue from predetermined speech samples transmitted by operator communication systems // Driving Assessment Conference. Vol. 5. University of Iowa. 2009.
4. Greeley H., Berg J., Friets E., *et al.* Fatigue estimation using voice analysis // Behavior Research Methods. 2007. Vol. 39, no. 3. P. 610–619. DOI: 10.3758/BF03193033.
5. openSMILE 3.0 - audEERING. Homepage. URL: <https://www.audeering.com/research/opensmile/A> (accessed: 15.11.2022).
6. Baykaner K., Huckvale M., Whiteley I., *et al.* The prediction of fatigue using speech as a biosignal // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) / ed. by V.K. Dediu A.-H. Martin-Vide C. 2015. Vol. 9449. P. 8–17. DOI: 10.1007/978-3-319-25789-1\_2.
7. Eyben F., Scherer K., Schuller B., *et al.* The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing // IEEE Transactions on Affective Computing. 2016. Vol. 7, no. 2. P. 190–202. DOI: 10.1109/TAFFC.2015.2457417.
8. Parada-Cabaleiro E., Costantini G., Batliner A., *et al.* DEMoS: an Italian emotional speech corpus: Elicitation methods, machine learning, and perception // Language Resources and Evaluation. 2020. Vol. 54, no. 2. P. 341–383. DOI: 10.1007/s10579-019-09450-y.
9. Freitag M., Amiriparian S., Pugachevskiy S., *et al.* auDeep: Unsupervised learning of representations from audio with deep recurrent neural networks // Journal of Machine Learning Research. 2018. Vol. 18. P. 1–5.
10. Hidalgo-Gadea G., Kreuder A., Krajewski J., Vorstius C. Towards better microsleep predictions in fatigued drivers: exploring benefits of personality traits and IQ // Ergonomics. 2021. Vol. 64, no. 6. P. 778–792. DOI: 10.1080/00140139.2021.1882707.
11. Fan X., Zhao C., Luo H., Zhang W. An event-related potential objective evaluation study of mental fatigue based on 2-back task // Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi. 2018. June. Vol. 35, issue 6. P. 837–844. DOI: 10.7507/1001-5515.201801064.
12. Trejo L., Kochavi R., Kubitz K., *et al.* Measures and models for predicting cognitive fatigue // Vol. 5797 / ed. by W.N. Caldwell J.A. 2005. P. 105–115. DOI: 10.1117/12.604286.
13. Yamada Y., Kobayashi M. Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults // Artificial Intelligence in Medicine. 2018. Vol. 91. P. 39–48. DOI: 10.1016/j.artmed.2018.06.005.
14. Matsumoto T., Ito K., Moritani T. The relationship between anaerobic threshold and electromyographic fatigue threshold in college women // European Journal of Applied Physiology and Occupational Physiology. 1991. Vol. 63, no. 1. P. 1–5. DOI: 10.1007/BF00760792.

15. Solberg G., Robstad B., Skjønberg O., Borchsenius F. Respiratory gas exchange indices for estimating the anaerobic threshold // Journal of Sports Science and Medicine. 2005. Vol. 4, no. 1. P. 29–36.
16. auDeep. URL: <https://github.com/auDeep/auDeep> (accessed: 15.11.2022).

Яковлев Александр Викторович, к.т.н., доцент, научно-исследовательский центр, ФГБВОУ ВО «Военно-медицинская академия имени С.М. Кирова» МО РФ (Санкт-Петербург, Российская Федерация), кафедра прикладной информатики, Санкт-Петербургский государственный университет аэрокосмического приборостроения (Санкт-Петербург, Российская Федерация)

Матыцин Вячеслав Олегович, к.м.н., научно-исследовательский центр, ФГБВОУ ВО «Военно-медицинская академия имени С.М. Кирова» МО РФ (Санкт-Петербург, Российская Федерация), ФГБОУ ВО ПСПбГМУ им. И.П. Павлова Минздрава России

Велюга Владислав Алексеевич, студент, Санкт-Петербургский государственный университет аэрокосмического приборостроения (Санкт-Петербург, Российская Федерация).

Найденова Ксения Александровна, к.т.н., с.н.с., научно-исследовательский центр, ФГБВОУ ВО «Военно-медицинская академия имени С.М. Кирова» МО РФ (Санкт-Петербург, Российская Федерация)

Пархоменко Владимир Андреевич, Высшая школа суперкомпьютерных систем и интеллектуальных технологий, Институт компьютерных наук и технологий, Санкт-Петербургский политехнический университет Петра Великого (Санкт-Петербург, Российская Федерация)

---

DOI: 10.14529/cmse220Y0Z

## PILOT EXPERIMENT TO CREATE A DATABASE OF RUSSIAN-LANGUAGE SPEECH DATA AND RECOGNIZE THE HUMAN FATIGUE STATE USING A NEURAL NETWORK

© 2022 A.V. Yakovlev<sup>1,2</sup>, V.O. Matytsin<sup>1,3</sup>, V.A. Velyuga<sup>2</sup>, X.A. Naidenova<sup>1</sup>,  
V.A. Parkhomenko<sup>4</sup>

<sup>1</sup>*Military Medical Academy (Akademika Lebedeva 6, St.Petersburg, 194044 Russia),*

<sup>2</sup>*Saint-Petersburg State University of Aerospace Instrumentation (Bolshaya Morskaya str. 67, St.Petersburg, 190000 Russia)*

<sup>3</sup>*FSBEI HE I.P. Pavlov SPbSMU MOH Russia (Lva Tolstogo str. 6-8, St.Petersburg, 197022 Russia),*

<sup>4</sup>*Peter the Great St.Petersburg Polytechnic University (Polytechnicheskaya str. 29, St.Petersburg, 195251 Russia)*

*E-mail: sven-7@mail.ru*

Received: 15.11.2022

Qualitative psychophysiological research is associated with the creation of accessible and well-organized databases that require a lot of preliminary work on the development of measuring complexes, including not only tools for measuring the psychophysiological parameters of a person, but also his emotional state, which is displayed in his facial expression, speech and behavioral patterns. Measuring systems should also include the means of processing experimental material. The essence of the study was to conduct a pilot experiment to create a prototype

of the Speech Data Base of Russian-speaking respondents, to obtain answers to some methodological questions that arise among specialists when using the database for the task of recognizing the state of human fatigue. A hardware and software complex has been developed that allows synchronously registering psychophysiological parameters, video recordings of behavioral reactions and audio recording of human speech. As a model of physical fatigue, a cardio-respiratory test with physical activity (load) was used. Before and after completing the test, volunteers read out a set of standard phonetically representative texts. The resulting audio recordings were processed using a specialized neural network capable of analyzing the integral spectral characteristics of sound. The resulting audio recordings were processed using a specialized neural network capable of analyzing the integral spectral characteristics of sound. The results of the pilot experiment showed the possibility of recognizing the state of fatigue of a person by his speech, which makes it possible to proceed to the creation of a large bank of audio recordings and the improvement of algorithms for recognizing the state of fatigue.

*Keywords: human fatigue recognition, speech database, instrumental complex, cardio-respiratory test, machine learning, neural network.*

## FOR CITATION

Yakovlev A.V., Matytsin V.O., Velyuga V.A., Naidenova X.A., Parkhomenko V.A. Pilot experiment to create a database of Russian-language speech data and recognize the human fatigue state using a neural network. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2022. Vol. X, no. Y. P. Z1–Z2. (in Russian) DOI: 10.14529/cmse220Y0Z.

*This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.*

## References

1. Zhang X.-J., Gu J.-H., Tao Z. Research of detecting fatigue from speech by PNN. Vol. 2. 2010. P. V2278–V2281. DOI: 10.1109/ICINA.2010.5636509.
2. Krajewski J., Batliner A., Golz M. Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. Behavior Research Methods. 2009. Vol. 41, no. 3. P. 795–804. DOI: 10.3758/BRM.41.3.795.
3. Krajewski J., Trutschel U., Golz M., *et al.* Estimating fatigue from predetermined speech samples transmitted by operator communication systems. Driving Assesment Conference. Vol. 5. University of Iowa. 2009.
4. Greeley H., Berg J., Friets E., *et al.* Fatigue estimation using voice analysis. Behavior Research Methods. 2007. Vol. 39, no. 3. P. 610–619. DOI: 10.3758/BF03193033.
5. openSMILE 3.0 - audEERING. Homepage. URL: <https://www.audeering.com/research/opensmile/A> (accessed: 15.11.2022).
6. Baykaner K., Huckvale M., Whiteley I., *et al.* The prediction of fatigue using speech as a biosignal. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) / ed. by V.K. Dediou A.-H. Martin-Vide C. 2015. Vol. 9449. P. 8–17. DOI: 10.1007/978-3-319-25789-1\_2.
7. Eyben F., Scherer K., Schuller B., *et al.* The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Transactions on Affective Computing. 2016. Vol. 7, no. 2. P. 190–202. DOI: 10.1109/TAFFC.2015.2457417.

8. Parada-Cabaleiro E., Costantini G., Batliner A., *et al.* DEMoS: an Italian emotional speech corpus: Elicitation methods, machine learning, and perception. *Language Resources and Evaluation*. 2020. Vol. 54, no. 2. P. 341–383. DOI: 10.1007/s10579-019-09450-y.
9. Freitag M., Amiriparian S., Pugachevskiy S., *et al.* auDeep: Unsupervised learning of representations from audio with deep recurrent neural networks. *Journal of Machine Learning Research*. 2018. Vol. 18. P. 1–5.
10. Hidalgo-Gadea G., Kreuder A., Krajewski J., Vorstius C. Towards better microsleep predictions in fatigued drivers: exploring benefits of personality traits and IQ. *Ergonomics*. 2021. Vol. 64, no. 6. P. 778–792. DOI: 10.1080/00140139.2021.1882707.
11. Fan X., Zhao C., Luo H., Zhang W. An event-related potential objective evaluation study of mental fatigue based on 2-back task. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*. 2018. June. Vol. 35, issue 6. P. 837–844. DOI: 10.7507/1001-5515.201801064.
12. Trejo L., Kochavi R., Kubitz K., *et al.* Measures and models for predicting cognitive fatigue. Vol. 5797 / ed. by W.N. Caldwell J.A. 2005. P. 105–115. DOI: 10.1117/12.604286.
13. Yamada Y., Kobayashi M. Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults. *Artificial Intelligence in Medicine*. 2018. Vol. 91. P. 39–48. DOI: 10.1016/j.artmed.2018.06.005.
14. Matsumoto T., Ito K., Moritani T. The relationship between anaerobic threshold and electromyographic fatigue threshold in college women. *European Journal of Applied Physiology and Occupational Physiology*. 1991. Vol. 63, no. 1. P. 1–5. DOI: 10.1007/BF00760792.
15. Solberg G., Robstad B., Skjøsberg O., Borchsenius F. Respiratory gas exchange indices for estimating the anaerobic threshold. *Journal of Sports Science and Medicine*. 2005. Vol. 4, no. 1. P. 29–36.
16. auDeep. URL: <https://github.com/auDeep/auDeep> (accessed: 15.11.2022).