

РАСПОЗНАВАНИЕ УТОМЛЕНИЯ ЧЕЛОВЕКА НА ОСНОВЕ АНАЛИЗА ЕГО РЕЧИ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ*

© 2023 А.В. Яковлев^{1,2}, В.О. Матыцин^{1,3}, В.А. Велюга²,
К.А. Найденова¹, В.А. Пархоменко⁴

¹Военно-медицинская академия им. С.М. Кирова
(194044 Санкт-Петербург, ул. Академика Лебедева, д. 6),

²Санкт-Петербургский государственный университет
аэрокосмического приборостроения

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

³Первый Санкт-Петербургский государственный медицинский университет
им. И.П. Павлова Минздрава России

(197022 Санкт-Петербург, ул. Льва Толстого, д. 6-8),

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

E-mail: sven-7@mail.ru, matitsin@list.ru, vladislav805@yandex.com,

ksennaidd@gmail.com, parhomenko.v@gmail.com

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 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 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, глубокая нейронная сеть.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Яковлев А.В., Матыцин В.О., Велюга В.А., Найденова К.А., Пархоменко В.А. Распознавание утомления человека на основе анализа его речи с помощью нейросетевых технологий // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2023. Т. 12, № 1. С. 46–60. DOI: 10.14529/cmse230103.

*Статья рекомендована к публикации программным комитетом Международной конференции «Data Analytics and Management in Data Intensive Domains — 2022».

Введение

Работа специалистов операторного профиля (далее — специалистов) 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 таких специалистов состояния утомления, что сопряжено с угрозой пропуска значимых сигналов и немотивированного реагирования на сигналы ложные. Поэтому контроль за развитием утомления у специалистов 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 специалиста;
- прогностическая ценность существующих математико-статистических моделей оценки работоспособности специалиста невелика в силу ограничений, накладываемых в большинстве случаев небольшими размерами изученных выборок.

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

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

Исследовательская активность, посвященная анализу речи человека сосредоточена в нескольких основных направлениях.

Первое направление сопряжено с совершенствованием алгоритмов обработки речевого сигнала. Это направление включает совершенствование программных инструментов для визуализации речевого сигнала и расчета его характерных признаков, что реализуется в частности в таких системах как PRAAT [2, 3], ISIP [4], openSmile [5]. К совершенствованию алгоритмов обработки речевого сигнала мы относим также работу по адаптации современных алгоритмов машинного обучения для решения задач анализа речи [1].

Второе направление связано с разработкой алгоритмов и теоретических подходов к распознаванию различных состояний и патологий человека на основе анализа его речи. К этому направлению можно отнести распознавание различных эмоциональных состояний, а также отклонений от нормальных психических и физиологических состояний.

Третье направление, особенно важное в области обеспечения безопасности труда специалистов операторского профиля деятельности, включает быструю и надежную оценку 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 (далее — БРД), составляющих основу машинного обучения. Трудоемкость этой проблемы 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 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 эксперимента соответственно. Основные выводы работы изложены в заключении.

1. Материалы и методы

Основу эксперимента составила синхронная регистрация речи в процессе чтения добровольцем стандартных фонетически представительных текстов. В 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 использовались два микрофона: профессиональный миниатюрный петличный микрофон AKG C 417^{III} и высокочувствительный метрологический микрофон PCB 378A14 совместно с усилителем PCB 482C. 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 auDeer [9]. Автоэнкодер реализован с помощью библиотеки TensorFlow версии 1.15. Обучение глубокой нейронной сети выполнялось на графической карте NVIDIA Quadro M4000.

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, размещенную на отдельном сервере в СУБД MySQL 5.8 и содержащую все собираемые данные [10].

Для удаленного управления экспериментом и доступа к таблицам и полям базы данных использовались следующие элементы комплекса: терминал оператора для удаленного управления показом текстов и записью речи добровольца; программа для записи речи добровольца по командам оператора; программа, выполняющая по командам оператора показ слайдов с текстом на проекторе для их прочтения добровольцем.

Сформированная база данных в целом, кроме сценариев и данных о добровольцах, содержит также описания классов состояний утомления и тексты, читаемые добровольцами. Описание каждого эксперимента включает: реализуемый сценарий, идентификатор добровольца, дату проведения эксперимента, речевые файлы.

Структура разработанного web-сервиса для доступа к таблицам и полям базы данных приведена на рис. 1.

3. Моделирование состояния утомления человека

В качестве моделей утомления обычно используются модель депривации сна [11], а также модель предъявления ментальной нагрузки, например тесты на переключение внимания либо решение арифметических задач в течение 100–180 минут [12–14]. 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, проводя 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 [15]. 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 данном этапе может считаться адекватной, при этом наиболее просто воспроизводимой.

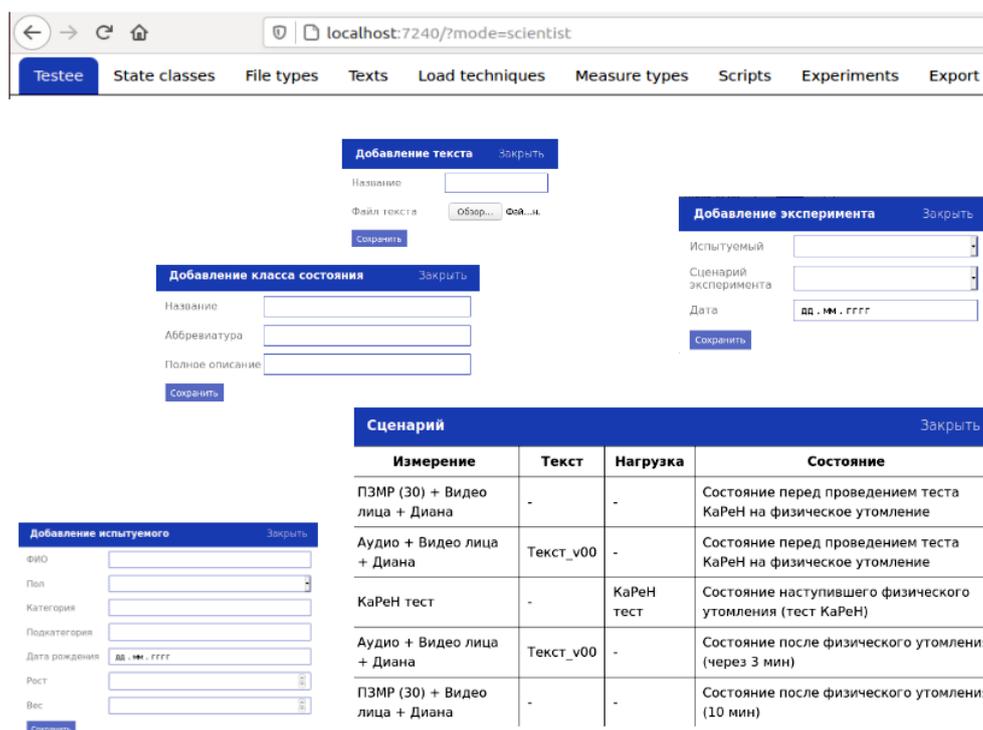


Рис. 1. Интерфейсы web-сервиса для доступа к базе данных

В качестве нагрузочного устройства в кардиореспираторном тесте (КаРен) использовали велоэргометр, поскольку он позволяет наиболее точно дозировать физическую нагрузку. Тест включает 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. Контролировалось достижение анаэробного порога (АП), критериями которого считаются следующие события: 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, начинает расти за счет гипервентиляции; достигается точка перекреста линий регрессии VO_2 и VCO_2 за счет роста VCO_2 [16].

Доброволец выполнял работу на велоэргометре до достижения им максимально переносимой физической нагрузки, при этом достижение анаэробного порога добровольцем рассценивалось как объективное подтверждение развития у него состояния физического утомления.

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 эксперимента регистрировали два состояния добровольца: состояние перед моделированием физического утомления (состояние «не утомлен», далее — S_1) и состояние после физического утомления (через 3 минуты) (состояние «утомлен», далее — S_2). Пе-

ред 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 тренировочной части и трех целевых частей:

- часть 1 («тренировочный текст») содержала несколько команд из ГОСТ 16600-72 для оценки средств связи (размер — 131 знак);
- часть 2 («командный текст») содержала большее количество команд из того же ГОСТ;
- часть 3 («проза») содержала фрагмент фонетически представительного текста;
- часть 4 («стих») содержала 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 исследования

Для решения задачи распознавания утомления человека была использована глубокая нейронная сеть с топологией автоэнкодера, реализованная в библиотеке auDeer [9]. 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 объем 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.

Последовательность применения рассматриваемой глубокой нейронной сети для задачи распознавания утомления человека по речи представлена на рис. 2 [17].

Она состоит из шести этапов:

1. Подготовка обучающего набора данных (англ. dataset) — образцов речевых сигналов с метками классов утомления (S_1 и S_2) для «работы» с нейронной сетью. Обычно такая подготовка состоит в «оформлении» этого набора данных в соответствии с требованиями парсера, который будет «разбирать» его на этапе извлечения спектрограмм. Это один из наиболее трудоемких для исследователя этапов, так как даже небольшое отклонение при оформлении датасета от требований парсера приводит к невозможности выполнения последующих этапов.
2. Извлечение спектрограмм: извлечение спектрограмм и данных о принадлежности этих спектрограмм к рассматриваемым классам из необработанных аудиофайлов.
3. Обучение автоэнкодера на извлеченных спектрограммах.
4. Генерация признаков обученной глубокой нейронной сетью.
5. Оценка сгенерированных признаков.
6. Экспорт сгенерированных признаков в форматы CSV или ARFF.

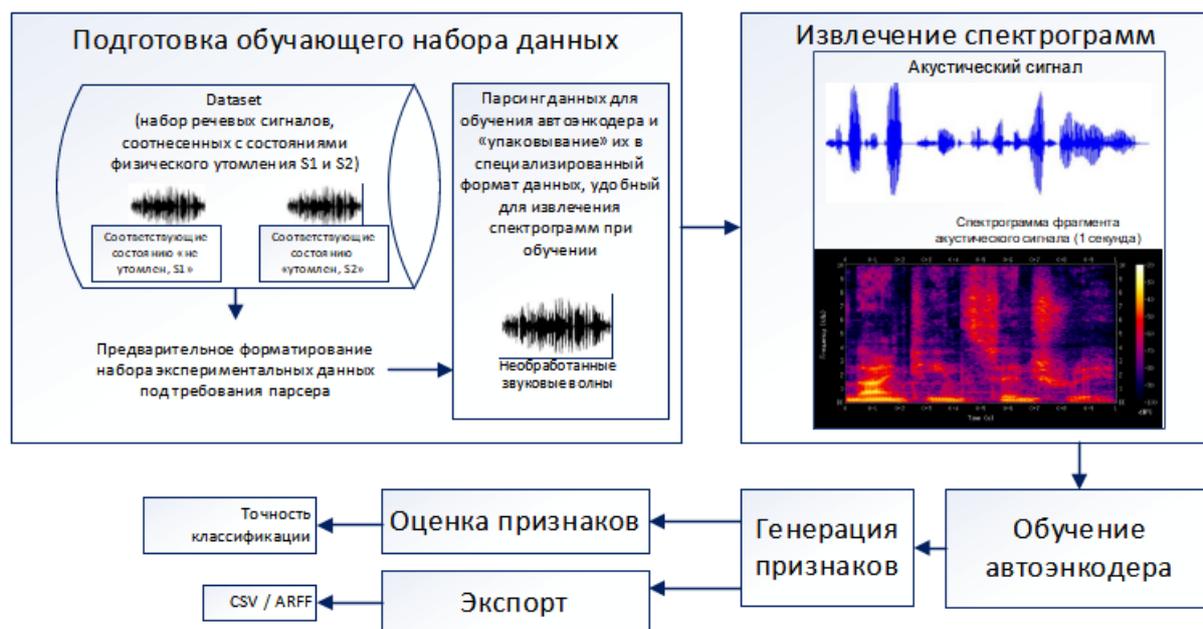


Рис. 2. Этапы использования библиотеки auDeep [9]

Для автоматизации процесса формирования датасетов в настоящей работе использовался web-сервис для доступа к базе данных (рис. 1, пункт меню «Export»). Пример результата «выгрузки» данных из БРД представлен на рис. 3.



Рис. 3. Структура выгружаемых комплексом данных

Подкаталог `data_set` содержит два подкаталога, соответствующие двум классам оцениваемого состояния утомления: подкаталог `001` содержит аудиофайлы, соответствующие состоянию добровольца «не утомлен», подкаталог `002` содержит аудиофайлы, соответствующие состоянию добровольца «утомлен». Каждый из подкаталогов содержал по 40 аудиозаписей. Таким образом, каждый из формируемых датасетов содержал 80 аудиозаписей.

В связи с тем, что целью настоящего исследования был поиск ответов на вопросы о предпочтительном типе текстов для чтения, о минимально достаточной длительности записываемых речевых фрагментов, о качестве используемых микрофонов, то варьировались датасеты, поступающие на вход нейронной сети.

Датасеты формировались исходя из значений следующих параметров:

- тип прочитанного текста, аудиозапись которого использовалась для обучения: «команда», «проза», «стихи», смешанный;
- длительность используемых для обучения нейронной сети фрагментов аудиозаписей (L): 7, 8, 9, 10 или 11 с.;
- канал регистрации: качественный петличный микрофон или высокочувствительный ненаправленный метрологический микрофон;
- уровень обрезания амплитуды записанных аудиоданных ниже указанного значения в децибелах: -45 дБ, -60 дБ, -75 дБ. Этим параметром устанавливался допустимый уровень «огрубления» исходных данных, с одной стороны, и, степень удаления из исходной записи низкоамплитудных шумов. Тем самым оценивалось, насколько используемый нейросетевой алгоритм чувствителен к шумам, присутствующих на исходной аудиозаписи.

Результаты оценки точности обучения нейронной сети — матрицы ошибок (англ. confusion matrix) для датасетов, сформированных с учетом вышеперечисленных параметров, приведены в табл. 1. Оценки качества классификации приведены в табл. 2.

Таблица 1. Матрицы ошибок обученных нейронных сетей для сформированных датасетов

Тип прочитанного текста	Уровень обрезания амплитуды аудиоданных ниже указанного значения												
	-45 дБ				-60 дБ				-75 дБ				
	Канал регистрации: качественный петличный микрофон												
	L, с.		S_1	S_2	L, с.		S_1	S_2	L, с.		S_1	S_2	
Команды	7	S_1	0.70	0.30	7	S_1	0.60	0.40	7	S_1	0.64	0.36	
		S_2	0.33	0.67		S_2	0.35	0.65		S_2	0.37	0.63	
Проза	9	S_1	0.68	0.32	9	S_1	0.69	0.31	9	S_1	0.66	0.34	
		S_2	0.27	0.73		S_2	0.21	0.79		S_2	0.32	0.68	
Стихи	10	S_1	0.65	0.35	7	S_1	0.54	0.46	8	S_1	0.63	0.37	
		S_2	0.26	0.74		S_2	0.37	0.63		S_2	0.35	0.65	
Смешанный	9	S_1	0.80	0.20	9	S_1	0.71	0.29	8	S_1	0.68	0.32	
		S_2	0.26	0.74		S_2	0.33	0.67		S_2	0.39	0.61	
		Канал регистрации: высокочувствительный ненаправленный метрологический микрофон											
		L, с.		S_1	S_2	L, с.		S_1	S_2	L, с.		S_1	S_2
Команды	7	S_1	0.66	0.34	7	S_1	0.60	0.40	7	S_1	0.68	0.32	
		S_2	0.33	0.67		S_2	0.31	0.69		S_2	0.44	0.56	
Проза	9	S_1	0.80	0.20	9	S_1	0.79	0.21	8	S_1	0.78	0.22	
		S_2	0.25	0.75		S_2	0.28	0.72		S_2	0.31	0.69	
Стихи	10	S_1	0.67	0.33	10	S_1	0.56	0.44	10	S_1	0.52	0.48	
		S_2	0.31	0.69		S_2	0.33	0.67		S_2	0.33	0.67	
Смешанный	9	S_1	0.73	0.27	8	S_1	0.83	0.17	8	S_1	0.80	0.20	
		S_2	0.29	0.71		S_2	0.39	0.61		S_2	0.43	0.57	

Таблица 2. Меры точности классификации обученных нейронных сетей для сформированных датасетов

Уровень обрезания амплитуды	Тип про- читанного текста	L , с.	Меры точности классификации			
			Accuracy	Sensitivity	Precision	F-мера
Канал регистрации: качественный петличный микрофон						
–45 дБ	Команды	7	0.685	0.690	0.67	0.680
	Проза	9	0.705	0.695	0.73	0.712
	Стихи	10	0.695	0.678	0.74	0.708
	Смешанный	9	0.770	0.787	0.74	0.762
–60 дБ	Команды	7	0.625	0.619	0.65	0.634
	Проза	9	0.74	0.718	0.79	0.752
	Стихи	7	0.585	0.577	0.63	0.602
	Смешанный	9	0.69	0.697	0.67	0.683
–75 дБ	Команды	7	0.635	0.636	0.63	0.633
	Проза	9	0.67	0.666	0.68	0.673
	Стихи	8	0.64	0.637	0.65	0.643
	Смешанный	8	0.645	0.655	0.61	0.632
Канал регистрации: высокочувствительный ненаправленный метрологический микрофон						
–45 дБ	Команды	7	0.665	0.663	0.67	0.666
	Проза	9	0.775	0.789	0.75	0.769
	Стихи	10	0.680	0.676	0.69	0.683
	Смешанный	9	0.72	0.724	0.71	0.717
–60 дБ	Команды	7	0.645	0.633	0.69	0.660
	Проза	9	0.755	0.774	0.72	0.746
	Стихи	10	0.615	0.603	0.67	0.635
	Смешанный	8	0.72	0.782	0.61	0.685
–75 дБ	Команды	7	0.620	0.636	0.56	0.595
	Проза	8	0.735	0.758	0.69	0.722
	Стихи	10	0.595	0.582	0.67	0.623
	Смешанный	8	0.685	0.740	0.57	0.644

5. Обсуждение результатов

Оценки качества распознавания целевого состояния S2 («утомлен») (табл. 1, 2) показывают, что применение нейронной сети с топологией автоэнкодер позволяет достичь точности распознавания целевого состояния до 75–79%. В зависимости от процедуры записи и характеристик записываемой речевой продукции эти результаты могут варьироваться.

Значимых отличий между использованием качественного петличного микрофона и высокочувствительного ненаправленного метрологического микрофона не выявлено. Это позволяет использовать в дальнейших исследованиях широко распространенные петличные микрофоны, но с внешней звуковой картой. Отдельным вопросом остается определение допустимых диапазонов технических характеристик используемого оборудования.

В связи с тем, что исследования проводились вне акустической безэховой камеры, а в офисном помещении в условиях относительной тишины, на аудиозаписях присутствовали шумы. Данные таблицы 2 иллюстрируют, что обрезание амплитуды исходного аудиосигнала ниже уровня -45 дБ (для аудиозаписей прозы ниже уровня -60 дБ) позволяет улучшить качество распознавания.

Одним из важных результатов исследования стало определение предпочтительного типа речевой продукции, регистрация которой позволялась распознавать нейронной сети состояние утомления с более высоким качеством. Данные таблицы 2 показывают, что использование петличного микрофона при обрезании амплитуды исходного аудиосигнала ниже уровня -45 дБ (то есть в условиях небольшого зашумления) позволяет распознавать состояние утомления со значениями F , равными 0.680 при обработке произносимых команд длительностью от 7 секунд, 0.712 — прозы длительностью от 9 секунд, 0.708 — стихотворений длительностью от 10 секунд и 0.762 для смешанного текста, состоящего из команд, прозы и стихотворений длительностью от 9 секунд. При меньшей длительности, вероятно, нейросети не хватает данных для того, чтобы найти признаки для различия состояний.

Таким образом, в результате исследования показана способность разработанной модели распознавать наступление состояния физического утомления у добровольцев по их речевой продукции.

Дальнейшее совершенствование разрабатываемой методики оценки состояния утомления по речи будет продолжено в направлении разработки БРД для исследования связи речевой продукции и умственного утомления, а также совершенствование рассмотренного в настоящей работе алгоритма для распознавания состояния 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 7–10 секунд.
4. Использование более чувствительного микрофона по сравнению с использованием петличного микрофона не дает заметной разницы качества распознавания состояния утомления.
5. Обрезание амплитуды исходного аудиосигнала ниже уровня -45 дБ (в ряде случаев ниже уровня -60 дБ) позволяет улучшить качество распознавания.

Литература

1. Zhang X.-J., Gu J.-H., Tao Z. Research of detecting fatigue from speech by PNN // 2010 International Conference on Information, Networking and Automation (ICINA). 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 // Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment 2009. University of Iowa, 2009. DOI: 10.17077/drivingassessment.1359.
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> (дата обращения: 15.11.2022).
6. Baykaner K., Huckvale M., Whiteley I., *et al.* The Prediction of Fatigue Using Speech as a Biosignal // Statistical Language and Speech Processing. Vol. 9449 / ed. by A.-H. Dediu, C. Martín-Vide, K. Vicsi. Cham: Springer, 2015. P. 8–17. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 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. URL: <http://jmlr.org/papers/v18/17-406.html>.
10. Яковлев А.В. Разработка распределенной программной системы для синхронизированного сбора речевых, видео- и психофизиологических данных о добровольце в процессе экспериментального исследования // Обработка, передача и защита информации в компьютерных системах '22: Сборник докладов второй международной научной конференции, Санкт-Петербург, Россия. Санкт-Петербург: Издательство Санкт-Петербургского государственного университета аэрокосмического приборостроения, 2022. С. 95–100.
11. 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.

12. Fan X., Zhao C., Luo H., Zhang W. An event-related potential objective evaluation study of mental fatigue based on 2-back task // *Journal of biomedical engineering*. 2018. Vol. 35, no. 6. P. 837–844. DOI: 10.7507/1001-5515.201801064.
13. Trejo L.J., Kochavi R., Kubitz K., *et al.* Measures and models for predicting cognitive fatigue // *Biomonitoring for Physiological and Cognitive Performance during Military Operations*. Vol. 5797 / ed. by J.A. Caldwell, N.J. Wesensten. International Society for Optics, Photonics. SPIE, 2005. P. 105–115. DOI: 10.1117/12.604286.
14. 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.
15. 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.
16. Solberg G., Robstad B., Skjønsberg 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. URL: <https://pubmed.ncbi.nlm.nih.gov/24431958/>.
17. Яковлев А.В. Использование многослойных сетей-автоэнкодеров для распознавания усталости человека на основе речевых данных // *Обработка, передача и защита информации в компьютерных системах '22: Сборник докладов второй международной научной конференции, Санкт-Петербург, Россия. Санкт-Петербург: Издательство Санкт-Петербургского государственного университета аэрокосмического приборостроения, 2022. С. 87–94.*

Яковлев Александр Викторович, к.т.н., доцент, научно-исследовательский центр, Военно-медицинская академия имени С.М. Кирова (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, Военно-медицинская академия имени С.М. Кирова (Санкт-Петербург, Российская Федерация), кафедра нормальной физиологии, Первый Санкт-Петербургский государственный медицинский университет им. И.П. Павлова Минздрава России (Санкт-Петербург, Российская Федерация)

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, Военно-медицинская академия имени С.М. Кирова (Санкт-Петербург, Российская Федерация)

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

RECOGNITION OF HUMAN FATIGUE BASED ON SPEECH ANALYSIS USING NEURAL NETWORK TECHNOLOGIES

© 2023 A.V. Yakovlev^{1,2}, V.O. Matytsin^{1,3}, V.A. Velyuga²,
X.A. Naidenova¹, V.A. Parkhomenko⁴

¹*S.M. Kirov Military Medical Academy*

(st. Akademika Lebedeva 6, St. Petersburg, 194044 Russia),

²*Saint-Petersburg State University of Aerospace Instrumentation*

(st. Bolshaya Morskaya 67, St. Petersburg, 190000 Russia),

³*Pavlov First Saint Petersburg State Medical University*

(st. Lva Tolstogo 6-8, St. Petersburg, 197022 Russia),

⁴*Peter the Great St. Petersburg Polytechnic University*

(st. Polytechnicheskaya 29, St. Petersburg, 195251 Russia)

E-mail: sven-7@mail.ru, matitsin@list.ru, vladislav805@yandex.com,

ksennaidd@gmail.com, parhomenko.v@gmail.com

Received: 15.11.2022

Qualitative psychophysiological research studies are 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 human, but also their emotional state, which is displayed in facial expression, speech and behavioral patterns. Measuring systems should also include the means of experimental material processing. The purpose of the study was to conduct an experiment on creating a prototype of the Speech Data Base of Russian-speaking respondents and to obtain answers to some methodological questions that arise among specialists when they use the database for the task of recognizing the state of human fatigue. A hardware and software complex has been developed that allows to synchronously register psychophysiological parameters, video recordings of behavioral reactions and audio recordings 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 obtained audio recordings were processed using a specialized neural network capable of analyzing the integral spectral characteristics of sound. The results of the experiment showed the possibility of recognizing the state of human fatigue based on speech analysis, 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, deep neural network.

FOR CITATION

Yakovlev A.V., Matytsin V.O., Velyuga V.A., Naidenova X.A., Parkhomenko V.A. Recognition of Human Fatigue Based on Speech Analysis Using Neural Network Technologies. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2023. Vol. 12, no. 1. P. 46–60. (in Russian) DOI: 10.14529/cmse230103.

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. 2010 International Conference on Information, Networking and Automation (ICINA). 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. Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment 2009. University of Iowa, 2009. DOI: 10.17077/drivingassessment.1359.
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. Statistical Language and Speech Processing. Vol. 9449 / ed. by A.-H. Dediu, C. Martín-Vide, K. Vicsi. Cham: Springer, 2015. P. 8–17. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 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. URL: <http://jmlr.org/papers/v18/17-406.html>.
10. Yakovlev A.V. Development of a distributed software system for synchronized collection of speech, video and psychophysiological data about a volunteer in the process of experimental research. Processing, Transmission and Protection of Information in Computer Systems '22: Proceedings of the Second International Scientific Conference, St. Petersburg, Russia. St. Petersburg: Publishing of the St. Petersburg State University of Aerospace Instrumentation, 2022. P. 95–100. (in Russian).
11. 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.
12. Fan X., Zhao C., Luo H., Zhang W. An event-related potential objective evaluation study of mental fatigue based on 2-back task. Journal of biomedical engineering. 2018. Vol. 35, no. 6. P. 837–844. DOI: 10.7507/1001-5515.201801064.

13. Trejo L.J., Kochavi R., Kubitz K., *et al.* Measures and models for predicting cognitive fatigue. *Biomonitoring for Physiological and Cognitive Performance during Military Operations*. Vol. 5797 / ed. by J.A. Caldwell, N.J. Wesensten. International Society for Optics, Photonics. SPIE, 2005. P. 105–115. DOI: 10.1117/12.604286.
14. 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.
15. 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.
16. 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. URL: <https://pubmed.ncbi.nlm.nih.gov/24431958/>.
17. Yakovlev A.V. The use of multilayer networks-autoencoders for the recognition of human fatigue on the basis of speech data. *Processing, Transmission and Protection of Information in Computer Systems '22: Proceedings of the Second International Scientific Conference, St. Petersburg, Russia*. St. Petersburg: Publishing of the St. Petersburg State University of Aerospace Instrumentation, 2022. P. 87–94. (in Russian).