DOI: 10.18720/SPBPU/2/id20-65

5. AUTOMATED SYSTEM FOR OBTAINING AND MINING DATA FOR MACHINE LEARNING IN EVALUATION OF THE SPECIALISTS' FUNCTIONAL STATE

Alexander Viktorovich Yakovlev, *cand. sci. (engineering), seniour researcher, Military Medical Academy, Russia, St.Petersburg, Akademika Lebedeva str., house 6, BOX 194044; associate professor, Saint-Petersburg State University of Aerospace Instrumentation, Russia, St.Petersburg, Bolshaya Morskaia str., house 67, BOX 190000. sven-7@mail.ru.*

Viacheslav Olegovich Matytsin, *cand. sci. (medicine), seniour researcher, Military Medical Academy, Russia, St.Petersburg, Akademika Lebedeva str., house 6, BOX 194044, matitsin@list.ru.*

Xenia Alexandrovna Naidenova, *cand. sci. (engineering), seniour researcher, Military Medical Academy, Russia, St.Petersburg, Akademika Lebedeva str., house 6, BOX 194044, ksennaidd@gmail.com.*

Vladimir Andreevich Parkhomenko, *software engineer, Peter the Great St. Petersburg Polytechnic University, Russia, St.Petersburg, Polytechnicheskaya str., house 29, BOX 195251, Vladimir.Parkhomenko@spbstu.ru.*

Annotation. *The most important stage in problems of diagnostics and forecasting of functional state and professional reliability of operators of complex man-machine systems is the correct collection and storage of diagnostic data. The authors propose and describe the solution of this problem in the form of a developed complex of multimodal registration of data on human functional state. The proposed solution uses technical and software tools to register large data sets of three types: audio, video and physiological data obtained from polygraph sensors. The complex produces primary data transformations, which allows to get a set of structured multimodal primary indicators, expressed in numerical form. Subsequently, the resulting array of multimodal indicators is used to search for regularities between them with the use of several libraries and machine learning frameworks.*

Keywords. *Operability, functional state of the human organism, hardware and software stand, audiovisual diagnostics, polygraph, face recognition, voice characteristics.*

АВТОМАТИЗИРОВАННАЯ СИСТЕМА СБОРА И ОБРАБОТКИ ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ ПРИ ОЦЕНКЕ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ СПЕЦИАЛИСТОВ

***Александр Викторович Яковлев**, к.т.н., ст. научный сотрудник, Военно-медицинская академия имени С.М.Кирова, Россия, Санкт-Петербург, ул. Академика Лебедева, д. 6, индекс 194044; доцент, Санкт-Петербургский государственный университет аэрокосмического приборостроения, Россия, Санкт-Петербург, ул. Большая Морская, д. 67, индекс 190000.*

***Вячеслав Олегович Матыцин**, к.м.н., ст. научный сотрудник, Военно-медицинская академия имени С.М.Кирова, Россия, Санкт-Петербург, ул. Академика Лебедева, д. 6, индекс 194044, matitsin@list.ru.*

***Ксения Александровна Найденова**, к.т.н., ст. научный сотрудник, Военно-медицинская академия имени С.М.Кирова, Россия, Санкт-Петербург, ул. Академика Лебедева, д. 6, индекс 194044, ksennaidd@gmail.com.*

***Владимир Андреевич Пархоменко**, программист, Санкт-Петербургский политехнический университет Петра Великого, Россия, Санкт-Петербург, ул. Политехническая, дом 29, индекс 195251, Vladimir.Parkhomenko@spbstu.ru.*

Аннотация. *Важнейшим этапом в задачах диагностики и прогнозирования функционального состояния и профессиональной надежности операторов сложных человеко-машинных систем является корректный сбор и хранение диагностических данных. Авторы предлагают и описывают решение данной проблемы в виде разработанного стенда мультимодальной регистрации данных о состоянии человека. В предлагаемом решении применяются технические и программные средства для регистрации больших массивов данных трех видов: аудиоданных, видеоданных и физиологических данных, получаемых с датчиков полиграфа. Стенд производит первичные преобразования данных, что позволяет получить набор структурированных мультимодальных первичных показателей, выраженных в численном виде. В последующем полученный массив мультимодальных показателей используется для поиска закономерностей между ними с помощью нескольких библиотек и фреймворков машинного обучения.*

Ключевые слова. *Работоспособность, функциональное состояние организма человека, аппаратно-программный стенд, аудиовизуальная диагностика, полиграф, распознавание лица, характеристики голоса.*

Введение

Существующие математические модели для оценки *функционального состояния организма* (далее — ФСО) специалиста и прогнозирования его работоспособности имеют недостаточно высокие прогностические возможности. Вместе с тем, современные условия профессиональной деятельности, улучшение технической оснащенности систем сбора данных о состоянии человека предъявляют повышенные требования к качеству прогностических моделей.

Приобретает все более широкое распространение анализ измерений, получаемых непосредственно с датчиков. Такие «сырые» сигналы, «снятые» непосредственно с модуля аналого-цифрового преобразования, обрабатываются в режиме реального времени и сравниваются с эталонными значениями, на основании чего выдается экспертное решение. Все большее значение приобретают методики оценки ФСО, основанные на регистрации аудио и видеосигналов, что связано с простотой их регистрации, их неинвазивностью и возможностью дистанционной записи. Таким образом, в современных условиях приобретает актуальность совершенствование способов оценки ФСО и прогнозирования работоспособности специалиста, а также совершенствование оборудования и программного обеспечения, предназначенных для этих целей.

5.1. Функциональное состояние организма и средства для его оценки

Под *ФСО* человека понимают совокупность характеристик его физиологических функций и психофизиологических качеств, которые несут наибольшую нагрузку в обеспечении профессиональной и поведенческой деятельности [5.24]. На рисунке 5.1 представлена взаимосвязь основных понятий, используемых в физиологии труда при решении задач по оценке ФСО специалиста и прогнозирования его работоспособности.

В работе предлагается методика синхронного сбора данных о физиологических реакциях организма и психоэмоциональных проявлениях человека,

Рис. 5.1. Взаимосвязь основных понятий физиологии труда при оценке функционального состояния организма специалиста

основанная на регистрации и сохранении данных с множества информационных каналов. Наглядно данный подход показан на рисунке 5.2.

Информационные каналы объединены в три модальности по принципу регистрации, сохранения и математической обработки данных, поскольку для сигналов каждой модальности будут использованы свои алгоритмы обработки.

Во-первых, это сигналы *физиологической* модальности, получаемые при многоканальной полиграфической регистрации физиологических реакций организма человека. Современный уровень развития средств измерений позволяет достоверно регистрировать и успешно использовать ряд биосигналов организма человека. Планируется регистрировать следующие биосигналы:

– фотоплетизматический сигнал для регистрации кровяного потока,
– изменение артериального давления,
– кожно-гальванический рефлекс,
– частота грудного и диафрагмального дыхания,
– изменение позы тела (смещение центра масс).

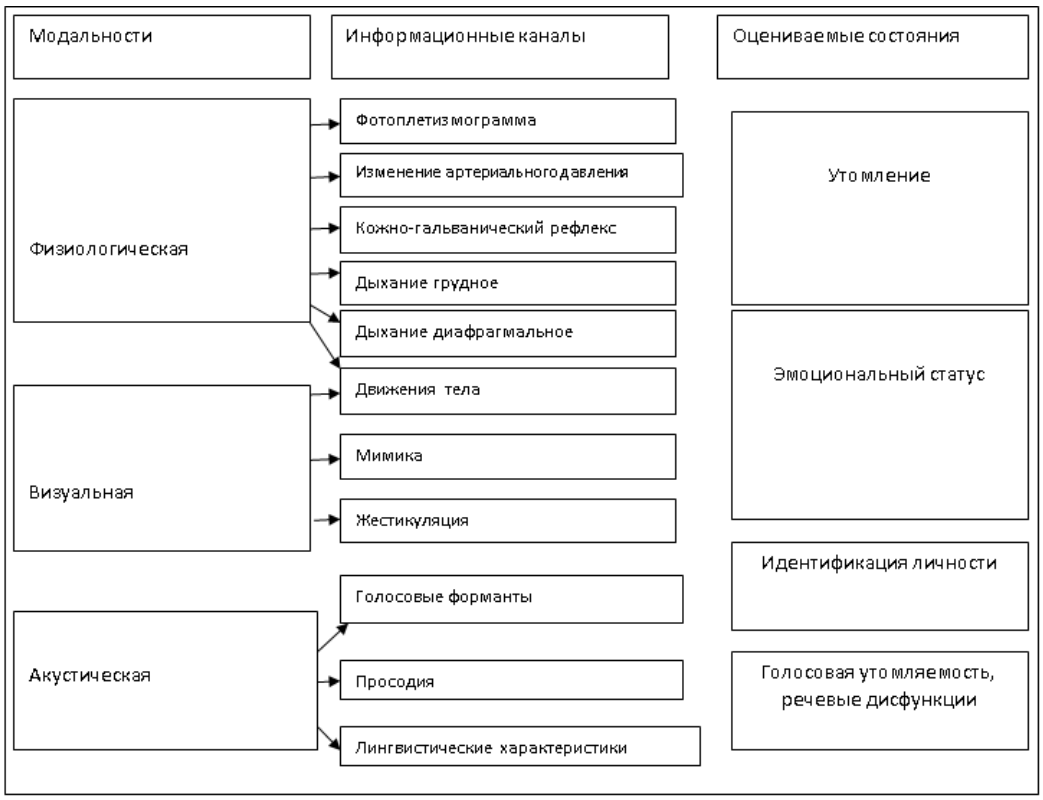

Рис. 5.2. Анализ сигналов трёх модальностей

Во-вторых, это сигналы *визуальной* модальности. Видеофиксация изображения лица человека, его мимики, позы и жестов с последующей их обработкой при помощи методов компьютерного зрения и машинного обучения позволяет выявить и формализовать основные классы психоэмоциональных проявлений человека в процессе его речевой деятельности.

В-третьих, это сигналы *акустической* модальности. Данная модальность связана с аудиорегистрацией речевого сигнала человека с одновременной регистрацией акустического шума, присутствующего в месте проведения исследования, с целью избежать систематической ошибки. Для регистрации речевого сигнала испытуемого предполагается использовать высокочувствительный микрофон, что позволит получить аудиозапись голоса максимально близкую к естественной.

Важнейшим этапом в задачах диагностики состояния человека является корректный сбор и хранение диагностических данных, поэтому актуальной

является задача создания аппаратно-программного средства, обеспечивающего подобный корректных сбор данных.

Был разработан аппаратно-программный стенд, позволяющий синхронно регистрировать физиологические, видео и акустические данные непосредственно со средств измерения и регистрации (далее — Стенд). Его архитектура представлена на рисунке 5.3 и повторяет типовую архитектуру систем бизнес-аналитики [5.8; 5.33]. Она содержит компоненты сбора данных, компоненты хранения данных и компоненты обработки данных [5.31].

В предлагаемом решении применяются технические и программные средства для регистрации больших массивов данных трех видов (т.е. трех модальностей): аудиоданных, видеоданных и данных, получаемых с физиологических датчиков [5.8; 5.26].

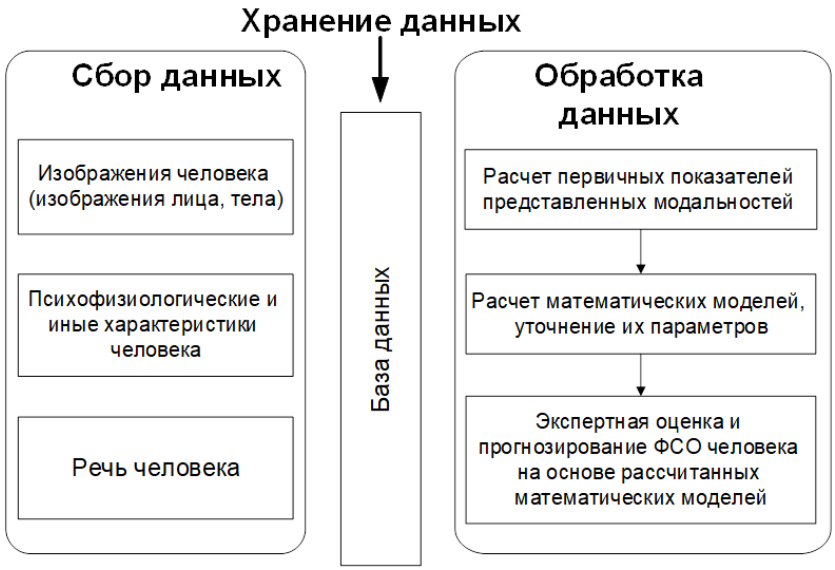

Рис. 5.3. Обобщенная структура Стенда [5.8; 5.33]

Компоненты сбора данных включают средства управления данными, определяющие порядок регистрации данных с оборудования и помещения их в базу данных. В базе данных (далее — БД) организуется и хранится для дальнейшего пользования вся собираемая информация, включая информацию о средствах измерения [5.34]. Обработка данных состоит в применении современных методов машинного обучения с целью извлечения знаний из собранных данных, например, представленных в работах [5.13; 5.14].

5.2. Сбор первичных данных

Источниками данных в предлагаемом Стенде являются следующие приборы:

– профессиональный компьютерный полиграф «Диана-7М»;
– видеокомплекс «Диана-Видео»;
– низкочастотный анализатор спектра UPV.

5.2.1. Профессиональный компьютерный полиграф

Профессиональный компьютерный полиграф «Диана-7М» (далее — ПКП) — один из самых совершенных российских полиграфов (детекторов лжи), адаптированный под научно-исследовательские задачи. Он регистрирует данные о состоянии человека с помощью набора датчиков для контактной регистрации состояния человека:

1. ВДХ — верхний датчик дыхания (пневматический), регистрирует грудное дыхание.
2. НДХ — нижний датчик дыхания (пневматический), регистрирует диафрагмальное дыхание человека.
3. КГР — датчик кожногальванической реакции.
4. АД — датчик артериального давления.
5. ПГ — датчик фотоплетизмограммы.
6. ТРМ — датчик двигательной активности.
7. Микрофонный датчик в виде петличного микрофона.

Датчики 1–6 подключаются к специализированному аналого-цифровому преобразователю (далее — АЦП), микрофонный датчик подключается к внешней звуковой карте M-Audio M-TrackMk.II. Таким образом, ПКП «Диана-7М» регистрирует 7 сигналов разной природы, которые с помощью специализированной библиотеки передаются пользователю в виде 6 сигналов и аудиофайла.

5.2.2. Видеокомплекс «Диана-Видео»

Видеокомплекс «Диана-Видео»[5.1] обеспечивает регистрацию невербального поведения человека с помощью видеокамер высокой чёткости (две IP видеокамеры Axis, разрешением 1280x720), а также качественную синхронизированную регистрацию его речи. IP видеокамеры комплекса «Диана-Видео» формируют поток кадров, которые передаются по сети.

5.2.3. Анализатор звуковых колебаний

Анализатор звуковых колебаний R&S UPV[5.2] с измерительными микрофонами 378A14 [5.3] регистрирует звуковые колебания в ближней зоне человека бесконтактным способом. В режиме цифрового магнитофона прибор может формировать два звуковых файла высокого качества. Управление прибором осуществляется по протоколу NI-VISA.

Важно отметить, что вышеперечисленные приборы регистрируют не сами показатели, а сигналы разных модальностей или, как их еще называют, «сырые» данные. Такие «сырые» данные не могут служить исходными данными для мультимодального анализа, как не прошедшие первичной обработки по выделению признаков и не имеющие четкой спецификации. В связи с этим возникла необходимость сохранения «сырых» данных в едином хранилище (БД) с целью последующей обработки. Задача была усложнена тем, что вышеперечисленные приборы имеют разные интерфейсы, разный формат представления данных, ограниченную емкость внутренней памяти и разные способы синхронизации.

Стенд построен по архитектуре «клиент-сервер», что позволяет преодолеть вышеуказанные сложности и обеспечить корректный синхронный сбор «сырых» данных с возможностью синхронного предъявления стимульного материала. На рисунке 5.4 представлена схема коммутации оборудования Стенда.

Помимо описанных выше источников данных, на рисунке 5.4 представлены коммутатор, сервер хранения и обработки данных и терминал управ-

[5.1] производитель «Поликониус», Россия

[5.2] производитель Rohde&Schwartz, Германия, № 48123-11 в Госреестре средств измерений

[5.3] производитель PCB Piezotronics, Inc., США, № 61720-15 в Госреестре средств измерений

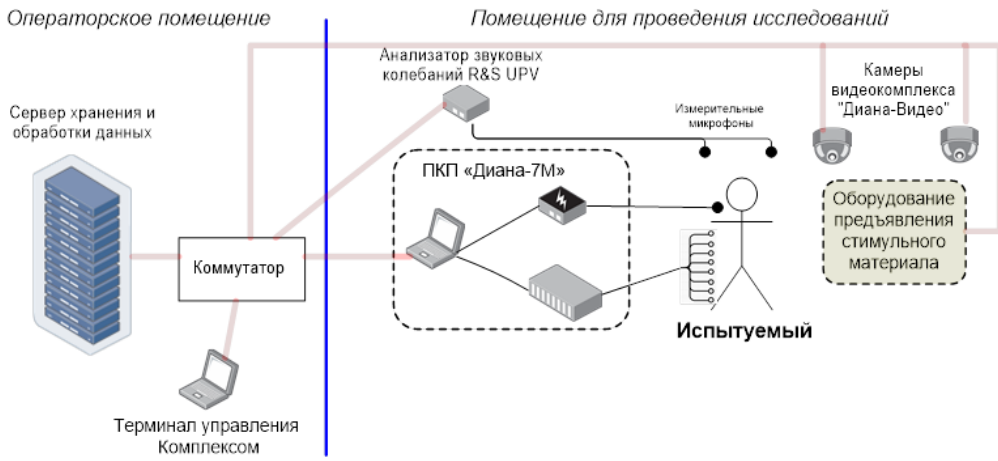

Рис. 5.4. Схема коммутации оборудования Стенда

ления Стендом. На рисунке 5.5 перечислены программные задачи (далее — ПЗ), реализованные в Стенде. Они условно сведены в три группы: обеспечивающие подготовку исследования, проведение исследования и обработку результатов исследования. Все ПЗ могут быть запущены с терминала управления Стендом.

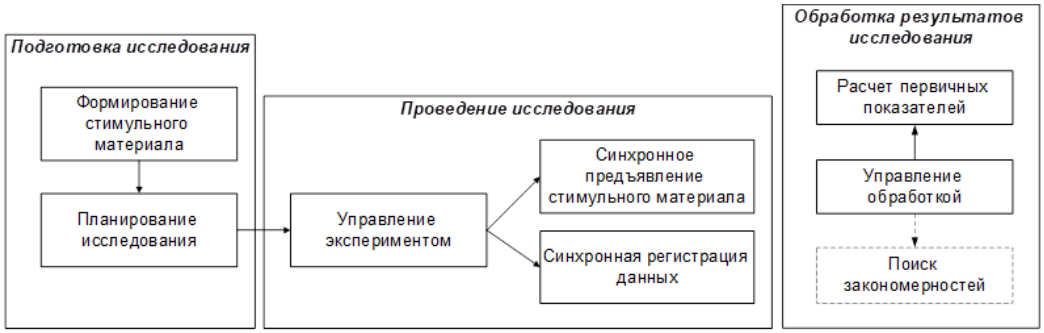

Рис. 5.5. Программные задачи, реализованные в Стенде

ПЗ «Формирование стимульного материала» предоставляет пользователю интерфейс, позволяющий внести в БД Стенда стимульный материал, который может быть предъявлен испытуемым в ходе проведения исследований. Стимульным материалом может быть видеоряд, аудиоряд, вопросники, тексты и т.п. Порядок и объем его предъявления испытуемым определяется при планировании конкретного исследования.

ПЗ «Планирование исследования» предоставляет пользователю интерфейс для занесения в БД Стенда сведений об испытуемом, о группах испытуемых, об используемом в исследовании стимульном материале и порядке его предъявления, о подключаемом регистрирующем оборудовании. Результатом выполнения этой ПЗ является план проведения исследования.

ПЗ «Управление экспериментом» предоставляет пользователю функционал по управлению процессом проведения эксперимента. Пользователь, находясь вне испытательного помещения, имеет возможность наблюдать в реальном времени процесс проведения исследования, включая отображение процесса регистрации показателей всех исследуемых модальностей и предъявления стимульного материала. В случае каких-либо отклонений от плана исследования пользователь может приостановить либо отменить эксперимент.

ПЗ «Синхронная регистрация данных» включает инициацию приборов, регистрирующих «сырые» данные, получение от них данных и передачу их по сети на сервер. ПЗ «Синхронное предъявление стимульного материала» обеспечивает синхронизированное с регистрацией данных предъявление испытуемому стимульного материала.

ПЗ «Расчет первичных показателей» производит преобразование «сырых» данных от регистрирующих приборов к форме, удобной для дальнейшего анализа, т.е. к первичным показателям. Ввиду того, что максимальное число первичных показателей может достигать нескольких тысяч по всем модальностям, ПЗ «Управление обработкой» предоставляет пользователю Стенда возможность (интерфейс) регулировать состав рассчитываемых первичных показателей и параметры их расчета. Полученные первичные показатели передаются в качестве исходных данных на фреймворки и библиотеки машинного обучения для поиска закономерностей в ПЗ «Поиск закономерностей». Эта ПЗ реализуется отдельно, не в составе Стенда, однако в тесной связи с ним.

5.3. Структура БД и подход к обработке данных

На программное обеспечение Стенда ложится задача не только сбора данных от регистрирующих и измерительных приборов, но и преобразование этих данных к форме, удобной для дальнейшего мультимодального анализа. Общий подход к обработке данных Стендом, который частично реализован, представлен на рисунке 5.6.

Рис. 5.6. Общий подход к обработке данных

Первоначально Стенд регистрирует массив данных, называемых «сырыми» или необработанными. Под ними подразумеваются данные, не прошедшие первичной обработки и не имеющие чёткой спецификации. Далее, в ходе преобразований этих «сырых» данных, с помощью специализированных программных библиотек рассчитываются вектора первичных показателей, пригодных для дальнейшей обработки средствами машинного обучения и поиска зависимостей между ними.

В базе данных Стенда хранится, помимо сырых данных, три массива показателей разной модальности: физиологические, акустические, визуальные.

Для их хранения на сервере использована открытая СУБД PostgreSQL. Структура базы данных Стенда представлена на рисунке 5.7.

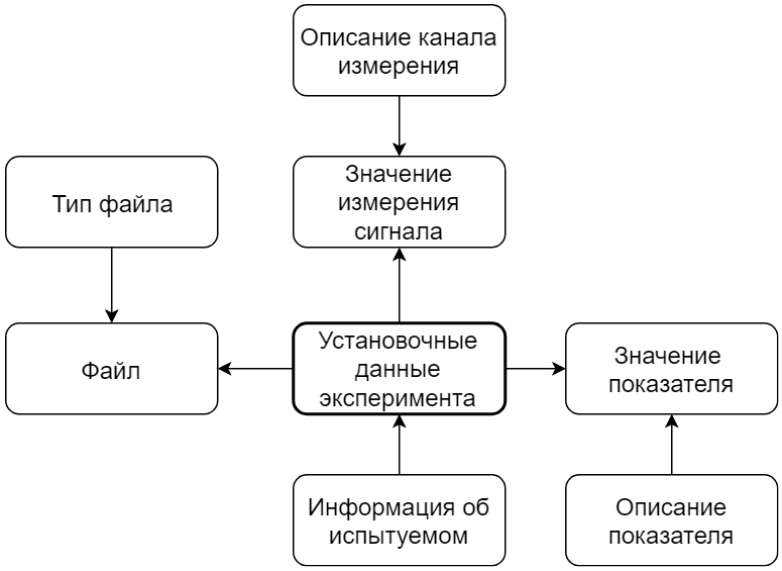

Рис. 5.7. Структура базы данных Стенда

Центральным звеном БД является таблица «Установочные данные эксперимента», связывающая конкретного испытуемого с его аудио- и видеозаписями, значениями измерений и вторичными показателями. Таблица «Значение измеряемого сигнала» хранит в себе значения того или иного показателя, полученные непосредственно в ходе эксперимента со ссылкой на тип (канал) измерения, а таблица «Значение показателя» — значения того или иного показателя, полученные после проведения эксперимента в ходе первичной и последующей обработок. Таблицы «Описание канала измерения» и «Описание показателя» используются для хранения общей информации о самом показателе или канале измерения. Полученной структуры БД достаточно для того, чтобы на данном этапе работы хранить результаты тех или иных экспериментов и результаты их обработки.

5.4. Обработка данных для оценки функционального состояния организма специалиста и прогнозирования его работоспособности

Обобщенная структура алгоритма оценки ФСО специалиста и прогнозирования его работоспособности дана на рисунке 5.8.

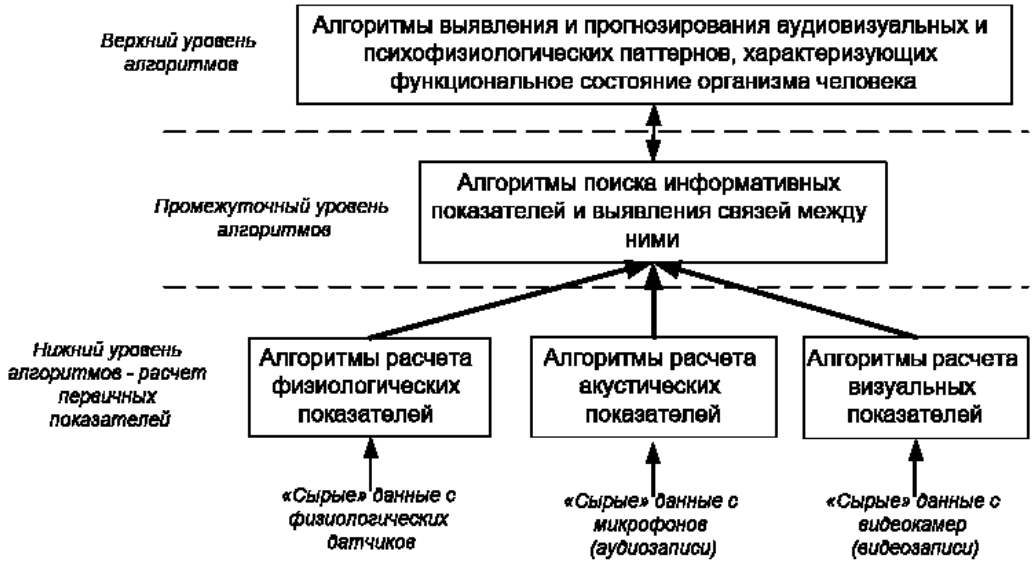

Рис. 5.8. Обобщенная структура алгоритма оценки ФСО специалиста и прогнозирования его работоспособности

5.4.1. Алгоритмы расчета физиологических показателей

К показателям физиологической модальности были отнесены шесть сигналов, регистрируемых ПКП «Диана-7М»:

A. грудного дыхания (от датчика ВДХ);
B. диафрагмального дыхания (от датчика НДХ);
C. кожно-гальванической реакции (от датчика КГР);
D. артериального давления (от датчика АД);
E. фотоплетизмограммы (от датчика ПГ);
F. двигательной активности (от датчика ТРМ).

Обработка вышеперечисленных сигналов с целью получения первичных показателей физиологической модальности производилась с помощью ма-

тематического пакета GNU Octave [5.16]. Данный пакет предоставляет улучшенный функционал для сложных математических вычислений, в т. ч. обработки сигналов. Такой выбор обусловлен отсутствием готового программного продукта, специализирующегося на обработке физиологических данных.

Рассмотрим обработку физиологических данных на примере обработки сигнала фотоплетизмографии с целью расчета ряда показателей, позволяющих оценить состояние сердечно-сосудистой системы в конкретный момент времени: частоту сердечных сокращений (ЧСС), частоту дыхания (ЧД), величину межударного интервала, и т. д.

Обратимся к теории. Метод фотоплетизмографии основан на измерении объемного пульса крови с помощью светового или инфракрасного датчика. С помощью фотоплетизмографии регистрируются три вида волн (рисунок 5.9) [5.17]:

A. Волны I-го порядка («пульсовые» волны) — синхронизированы с колебаниями сердечной мышцы.
B. Волны II-го порядка («дыхательные» волны) — синхронизированы с частотой дыхания. Их амплитуда и период обычно меньше таковых у волн I-го порядка.
C. Волны III-го порядка — отражают периодичность в активности сосудодвигательного центра, хотя часто они носят апериодический характер.

Таким образом, для измерения ЧСС с помощью плетизмограммы необходимо выделить из неё пульсовые волны и измерить интервалы между их пиками. Рассмотрим необходимые преобразования. Для измерения ЧСС плетизмограмма должна пройти предобработку, состоящую из нескольких этапов.

A. «Выделение дыхательных волн». Дыхательные волны имеют меньшую периодичность, чем пульсовые волны, поэтому для их выделения использовалась полиномиально-регрессионная модель [5.19], применяемая на всей выборке или некотором её участке. Из-за разницы в периодах волн, пульсовые волны распознавались регрессионной моделью как шум, а дыхательные — как необходимый сигнал.
B. «Сглаживание помех дыхательных волн». Так как регрессионная модель применяется по принципу скользящего окна (некоторого интер-

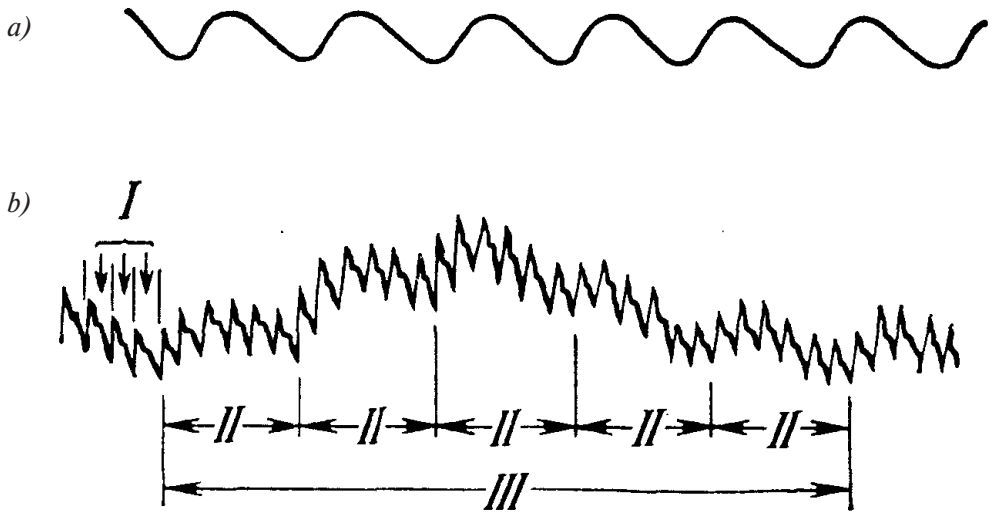

Рис. 5.9. Синхронно зарегистрированные показатели: *a* — пневмограмма; *b* — пальцевая плетизмограмма [5.17]

вала), на границах окна полученный ряд имеет нежелательные помехи, способные, при выделении пульсовых волн, исказить их. Указанные помехи были сглажены с помощью медианной фильтрации [5.11]. Медианная фильтрация дала хороший результат и минимизировала помехи.

C. «Выделение полезного сигнала». Была произведена операция вычитания выделенного ряда, соответствующего дыхательной волне, из исходного сигнала, в результате чего был получен сигнал пульсовой волны. Далее, для лучшего распознавания пиков пульсовых волн, полученный сигнал был усилен на несколько децибел.

D. «Расчет показателей». Производится поиск пиков пульсовых волн путем выделения всех участков ряда с амплитудой выше 10 и нахождения их локальных экстремумов, после которого координаты экстремумов сопоставляются с рядом временных меток и находятся их интервалы. Это позволяет рассчитать значение ЧСС для произвольного момента времени. После нахождения значений ЧСС на всем рабочем участке, они помещаются в БД с указанием временных меток. Также с помощью дискретного преобразования Фурье (ДПФ) возможно получить частотный спектр межударного интервала, который можно проанали-

зировать и получить таким образом набор частотных характеристик междударного интервала.

Обработка физиологических сигналов грудного и диафрагмального дыхания подразумевает оценку частоты дыхания, а также получение соотношения амплитуд каждого из приведённых типов дыхания с целью установить, какой из них преобладает в конкретный момент времени.

Оценка частоты дыхания испытуемого с помощью данных сигналов сводится к обнаружению пиков (точек экстремума), как и в случае с фотоплетизмограммой (рисунок 5.10). После этого производится подсчет пиков за одну минуту, что позволяет получить частоту дыхания испытуемого (рисунок 5.11). Преобладание грудного или диафрагмального дыхания устанавливается путем вычета сигнала НДХ из сигнала ВДХ. Если значение выше нуля — преобладает грудное дыхание, в противном случае — диафрагмальное (рисунок 5.12).

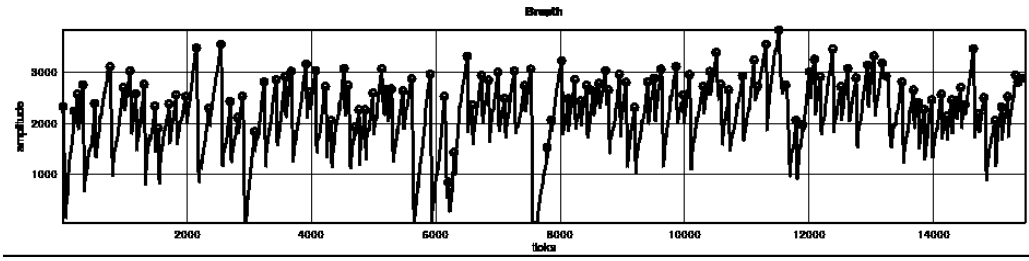

Рис. 5.10. Обнаруженные пики сигнала грудного дыхания, обозначаюшие верхнюю точку вдоха

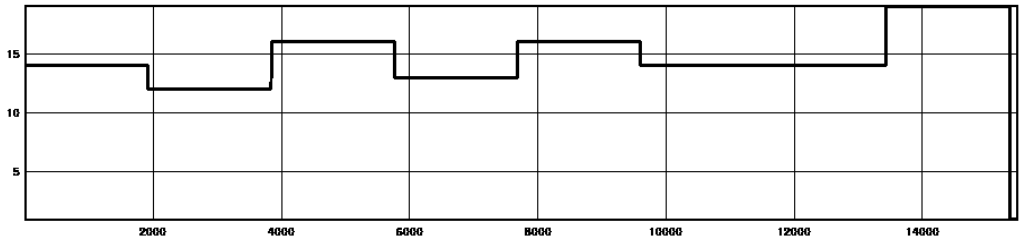

Рис. 5.11. Зарегистрированная частота дыхания испытуемого в ходе эксперимента

Аналогичным образом, с применением ряда математических преобразований, были рассчитаны другие показатели физиологической модальности (таблица 5.1).

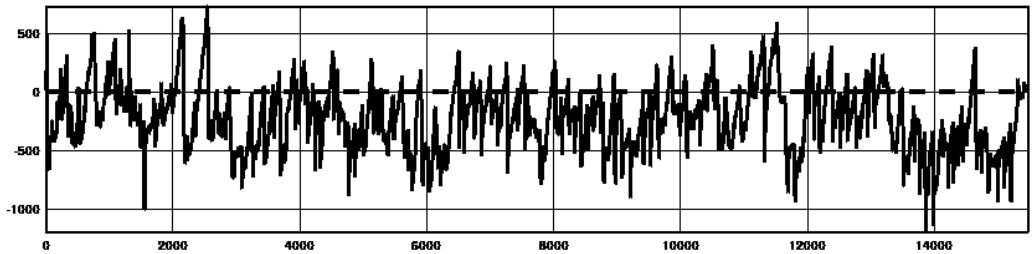

Рис. 5.12. График соотношений амплитуд грудного и диафрагмального дыхания

Таблица 5.1

Показатели физиологической модальности

Исходный сигнал	Рассчитанные первичные показатели
Сигнал ФПГ	Частота сердечных сокращений (далее — ЧСС) Междударный интервал (далее — МИ) Вариабельность сердечного ритма (далее — ВСР)
Сигнал ВДХ	Частота грудного дыхания, Амплитуда грудного дыхания
Сигнал НДХ	Частота диафрагмального дыхания Амплитуда диафрагмального дыхания
Сигнал КГР	Уровень долгосрочной реакции КГР Уровень краткосрочной реакции КГР
Сигнал ТРМ	Частота тремора

5.4.2. Алгоритмы расчета визуальных показателей

Для цифровой обработки полученных видеоизображений использовалась библиотека OpenFace 2.0 [5.1]. Библиотека разбивает входящий видеопоток на отдельные кадры (статические изображения) и анализирует их по отдельности, проводя ряд последовательных преобразований статического изображения в соответствии с алгоритмом активных моделей внешнего вида[5.4] [5.9]. Согласно ему для разметки ключевых лицевых точек статические модели изображений «подгоняются» под реальное изображение (рисунки 5.13a, 5.13b). Результатом являются координаты 68 точек, которые описывают контур лица, носа, глаз, бровей и губ (рисунок 5.13c).

[5.4]англ. Active Appearance Models, AAM

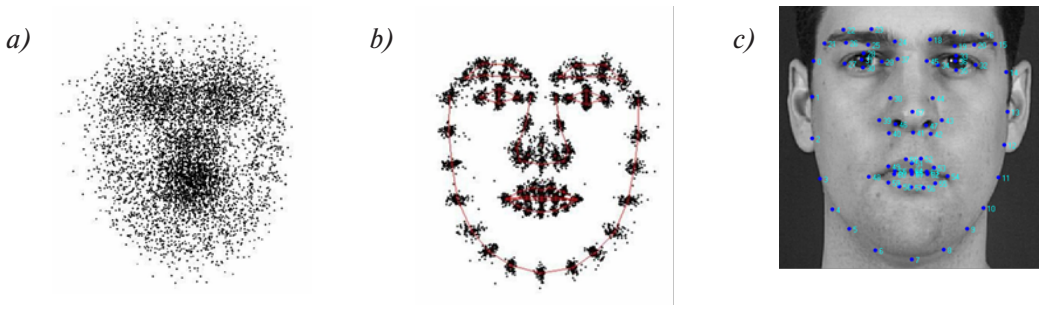

Рис. 5.13. Принцип работы алгоритма AAM [5.9], в том числе: *a* — невыровненная форма; *b* — выровненная форма; *c* — результат

Функционал библиотеки позволяет производить захват лица человека и расчёт базовых анатомических точек на его лице, определять наклон и поворот лица в пространстве и распознавать лицевые движения. По каждому кадру библиотека рассчитывает 714 показателей, структура которых представлена на рисунке 5.14. Однако с точки зрения возможной связи этих показателей с параметрами физиологической модальности интерес представляет только их часть (обведены эллипсами на рисунке 5.14).

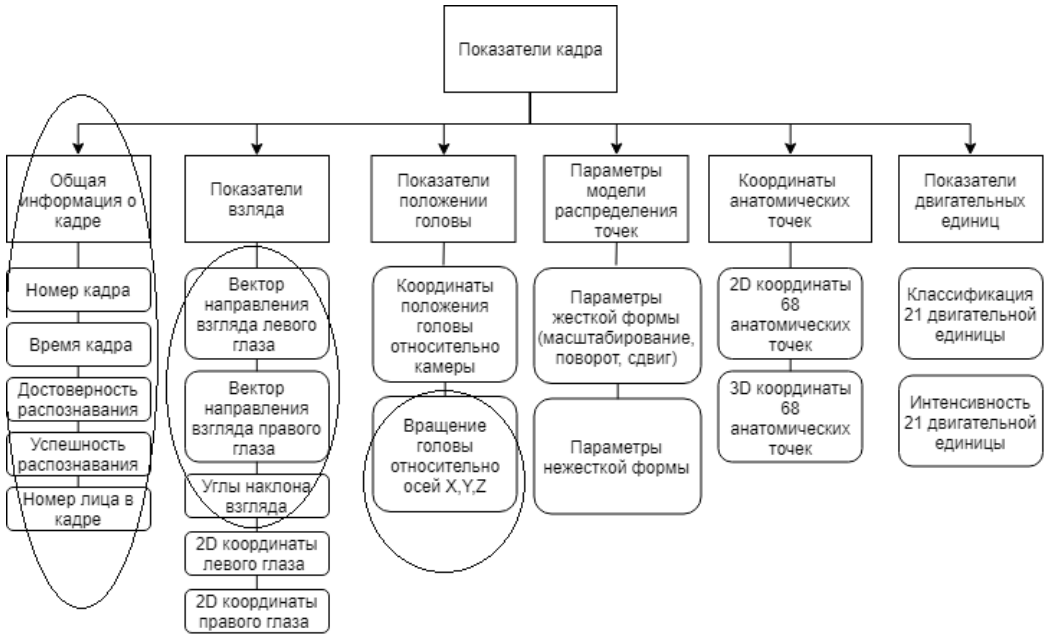

Рис. 5.14. Структура показателей обработанного кадра

В частности, широко используются в мире, иногда в сочетании с анализом речи, в диагностике нейропсихиатрических расстройств, биполярных расстройств, депрессии и шизофрении [5.4; 5.5; 5.7] показатели двигательных единиц[5.5]. Их родоначальник, Пол Экманн, разработавший «Систему кодирования лицевых движений»[5.6] [5.6]. Рассчитываемая библиотекой OpenFace 2.0 группа показателей двигательных единиц содержит классификацию и интенсивность 21 двигательной единицы. На рисунке 5.15 представлен пример их расчета для произвольного изображения.

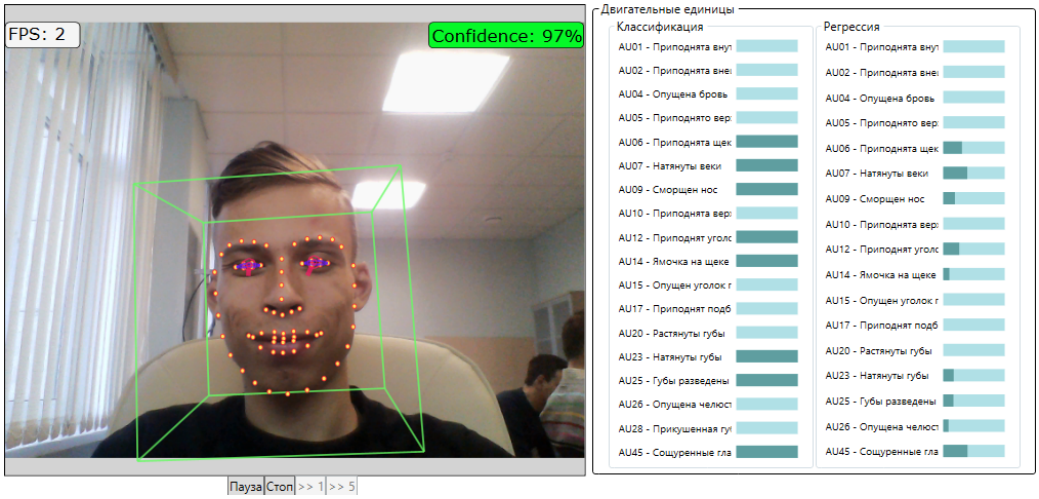

Рис. 5.15. Пример работы библиотеки OpenFace 2.0 в составе Стенда

Наличие кодов двигательных единиц дает возможность автоматически оценить основные эмоции на лице человека на основе вышеупомянутой «Системы кодирования лицевых движений» в соответствии с формулой:

$$E = F \cdot I, \tag{5.1}$$

где E — вектор эмоций, F — вектор кодов двигательных единиц, I — интенсивность двигательных единиц. Элементы вектора I могут принимать значения:

- A — слабо различимое;
- B — незначительное;
- C — заметное или ярко выраженное;

[5.5] англ. action units, AU

[5.6] англ. Facial Action Coding System, FACS

– D — сильное или крайне заметное;
– E — предельное;
– $*$ — может быть любой интенсивности.

Пример 5.1. Эмоция «Удивление» кодируется «1 + 2 + 5B + 26», эмоция «Страх» — «1 + 2 + 4 + 5* + 20* + 25,26» или «1 + 2 + 4 + 5* + 25,26».

Для выделения пульсаций на видеоизображении применялся алгоритм Эйлерова усиления видео, который выполняет усиление цвета или движения на видео. Для выделения пульсаций использовалось усиление цвета, благодаря чему предоставляется возможность расчета пульса человека.

Результаты предварительных экспериментов показывают эффективность фотоплетизмографии по видеоизображению лица человека с использованием Эйлерова усиления видео. Для практического использования системы необходимо выделять лицо человека при различных вариантах регистрации видеоизображения и устранять различные внешние факторы, например изменение освещенности лица при видеосъемке.

5.4.3. Алгоритмы расчета акустических показателей

Для параметрического представления речевого (акустического) сигнала он делится на короткие сегменты, и каждый сегмент преобразуется в вектор признаков. В результате входной сигнал представляется последовательностью векторов признаков. В основе такого подхода лежит допущение, что речь может рассматриваться как стационарная на этих коротких интервалах. Для более точного описания сигнала речевые сегменты берутся с перекрытием. Процесс создания речевых сегментов выполняется с помощью метода окна, т.е. путем перемножения сигнала с некоторой функцией окна для того, чтобы разрывы на границах окна были ослаблены. Обычно для этих целей используется окно Хэмминга [5.10].

Для параметрического представления речевого сигнала, т.е. получения первичных показателей речевого сигнала, была использована библиотека openSMILE [5.21], позволяющая рассчитать большой набор показателей речевого сигнала. Однако, в Стенде использовались не все показатели, а только их часть, описанная в качестве конфигурационных файлов для библиотеки openSMILE (таблица 5.2). На рисунке 5.16 представлена структура рассчиты-

ваемых показателей или, как их еще называют, низкоуровневых дескрипторов[5.7] [5.30].

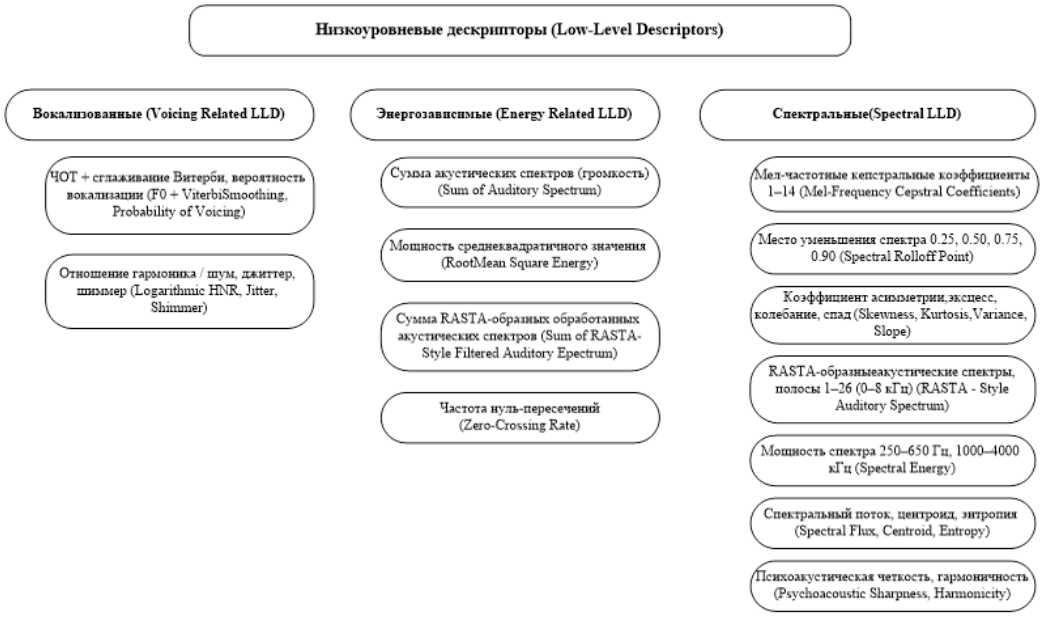

Рис. 5.16. Группы низкоуровневых дескрипторов, рассчитываемых в openSMILE

Спектральная группа низкоуровневых дескрипторов основана на преобразовании Фурье. В частности, мел-частотные кепстральные коэффициенты и RASTA-образные акустические спектры используются в автоматическом распознавании речи и в поиске информации о музыке [5.18]. Спектральный центроид является одним из наиболее часто используемых параметров тембра [5.27]. К *энергозависимой* группе относятся дескрипторы, описывающие просодические, амплитудные характеристики речи. Так, суммы RASTA-образных обработанных акустических спектров определяют состав шума в зависимости от частоты [5.20]. *Вокализированная* группа дескрипторов описывает голосовые характеристики и качество звука. Так, джиттер и шиммер используются не только в анализе голосовой патологии, но также являются хорошими дескрипторами общего качества звука. Частота основного тона определяет основные частотные характеристики акустических данных, такие как тон голоса человека [5.18]. Процедура запуска и работа с программ-

[5.7] англ. Low Level Descriptors, LLD

ной средой OpenSmile обеспечивают расчет вышеперечисленных групп низкоуровневых акустических дескрипторов (таблица 5.2)

Таблица 5.2

Группы признаков, рассчитываемые с помощью конфигурационных файлов

Рассчитываемая группа признаков	Краткое описание
Мел-частотные кепстральные коэффициенты 1–14 (Mel-frequency Cepstral Coefficients, MFCC features)	Размер кадра устанавливается равным 25 мс со скоростью 10 мс. Частотный диапазон Мел-спектра устанавливается от 0 до 8 кГц.
Кепстральные коэффициенты для перцептивного линейного предсказания (Perceptual Linear Prediction Cepstral Coefficients, PLP-CC)	Используются для задач фонетической сегментации [5.29]. Размер кадра устанавливается равным 25 мс со скоростью 10 мс.
Просодические признаки (Prosodic features)	Частота основного тона (F0), вероятность голоса, контуры громкости
The INTERSPEECH 2009 Emotion Challenge feature set	Набор функций, утвержденный конференцией The INTERSPEECH 2009 Emotion Challenge, для распознавания и исследования эмоциональной окраски речи. Набор содержит 384 функции в качестве статистических функционалов, применяемых к низкоуровневым контурам дескриптора. Функции сохраняются в формате Arff (WEKA), благодаря чему новые экземпляры добавляются к существующим (используется для пакетной обработки, когда openSMILE неоднократно вызывается для извлечения функций из нескольких файлов в один файл функций).

Продолжение таблицы 5.2

Рассчитываемая группа признаков	Краткое описание
The INTERSPEECH 2010 Paralinguistic Challenge feature set	Набор функций, утвержденный конференцией The INTERSPEECH 2010 Paralinguistic Challenge, для распознавания и исследования паралингвистических особенностей речи. Основные предназначения пакета направлены на определения пола, возраста и уровня заинтересованности.
The INTERSPEECH 2011 Speaker State Challenge feature set	Набор функций, утвержденный конференцией The INTERSPEECH 2011 Speaker State Challenge, для распознавания и исследования уже отдельного типа задач: задача распознавания алкогольной интоксикации и задача распознавания сонливости по человеческой речи.
The INTERSPEECH 2012 Speaker Trait Challenge feature set	Набор функций, утвержденный конференцией The INTERSPEECH 2012 Speaker Trait Challenge feature, для распознавания и исследования следующего типа задач: вычислительный анализ личности, симпатия и патологии в речи.
The INTERSPEECH 2013 ComParE Challenge feature set	Набор функций The INTERSPEECH 2013 ComParE Challenge впервые предоставляет единый тестовый стенд для социальных сигналов, таких как смех в речи. С помощью данного набора рассматривается аутизм и его проявления в речи.
The MediaEval 2012 TUM feature set for violent video scenes detection	Набор функций The MediaEval 2012 TUM для распознавания и исследования сцен насилия в фильмах.
The openSMILE/openEAR ‘emobase’ set	Старый базовый набор (emobase) от openEAR, состоящий из 988 акустических функций для распознавания эмоций.
The large openSMILE emotion feature set	Пакет для извлечения большего набора (emo_large) функций с большим количеством функций и включенным LLD (всего 6 552 функции).

Продолжение таблицы 5.2

Рассчитываемая группа признаков	Краткое описание
The openSMILE ‘emobase2010’ reference set	Набор функций (emobase2010), основанный на INTERSPEECH 2010 Paralinguistic Challenge feature set.
Audio-visual features based on INTERSPEECH 2010 audio features	Папка config / audiovisual содержит два файла конфигурации для функций видео (video.conf) и синхронизированное извлечение аудиовизуальных функций (audiovideo.conf). Аудио функции и набор функционалов, которые применяются как к аудио, так и к видео низкоуровневым функциям, взяты из набора функций INTERSPEECH 2010 Paralinguistic Challenge

5.4.4. Подходы к выявлению аудиовизуальных и психофизиологических паттернов, характеризующих ФСО специалиста

Для оценки характеристик БД и возможностей Стенда был проведен пилотный эксперимент [5.32]. Пятнадцати испытуемым была предъявлена речевая нагрузка в виде чтения специально подготовленных текстов разного уровня сложности. В ходе чтения текстов была осуществлена синхронная регистрация аудиовизуальных и психофизиологических показателей. Итоговый массив «сырых» данных составил 1,6 ГБ, из них: физиологические данные объемом 16 МБ, акустические — 1,2 ГБ, видео — 440 МБ. Такой значительный объем памяти, занимаемой первичными показателями, связан с тем, что по каждому показателю в ходе 10-минутного эксперимента с одним испытуемым регистрируется около 15000 измерений. В связи с этим стал актуальным вопрос понижения размерности данных, а также поиска алгоритма выявления закономерностей в данных, представленных временными рядами. На рисунке 5.17 представлен пример таких данных.

Поиск зависимостей в указанном наборе показателей является сложной задачей, в первую очередь связанной с анализом временных рядов и широким применением современных фреймворков машинного обучения. Для оптимизации поиска паттернов и закономерностей в вышеописанном наборе

данных авторами был проведен анализ современных фреймворков и библиотек машинного обучения с целью определения множества наиболее подходящих для стоящей перед ними задачи. Используемый набор инструментов машинного обучения должен удовлетворять следующим требованиям:

- возможность работы со слабо структурированными данными большой размерности (так называемыми «большими данными»);
- возможность работы в режиме «обучения без учителя»;
- поддержка многопоточной параллельной обработки данных, позволяющая максимально использовать процессорные мощности и ускорить процесс обработки;
- поддержка серверных операционных систем, в частности, ОС Ubuntu Server;
- возможность взаимодействия с базой данных через СУБД PostgreSQL;
- свободная модель распространения, предполагающая его бесплатность и открытый исходный код.

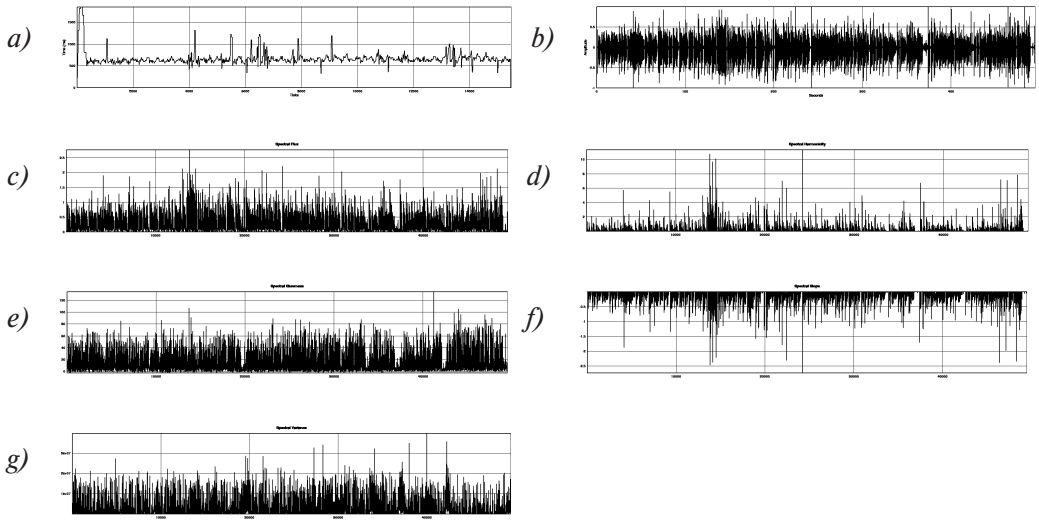

Рис. 5.17. Пример сочетанного изменения ЧСС обследуемого и акустических характеристик его речи: *a* — ЧСС; *b* — акустический сигнал; *c* — спектральный поток; *d* — гармоничность спектра акустического сигнала; *e* — асимметрия спектра акустического сигнала; *f* — наклон спектра акустического сигнала; *g* — дисперсия спектра акустического сигнала

С учетом вышеперечисленных ограничений, авторами для дальнейшего рассмотрения и возможного использования были выбраны следующие библиотеки и фреймворки машинного обучения, успешно работающие с теми типами данных, поиск закономерностей среди которых является целью настоящего исследования. Рассмотрим эти библиотеки подробнее.

TensorFlow [5.25] — фреймворк компании Google для решения задач построения и тренировки нейронной сети с целью автоматического нахождения и классификации образов. Задача TensorFlow — обучать нейросети обнаруживать и распознавать паттерны и корреляции в массивах данных. API этой библиотеки реализовано для Python, также существуют реализации для C++, Haskell, Java и Go. В библиотеке реализованы сверточные нейронные сети (CNN), рекуррентные сети и сети с памятью (RNN). Библиотека поддерживает ускорение с помощью графических процессоров, используя технологию CUDA. Эта библиотека активно используется для распознавания объектов на видео и классификации данных.

Библиотека Scikit-Learn [5.22] создана на основе математических пакетов Python: NumPy, SciPy, и matplotlib. Она реализует обучение с учителем и без, решает задачи классификации, кластеризации, сокращение размерности и др. [5.23].

Библиотека Theano [5.28] написана на Python и является открытым аналогом пакетов Mathematica и MATLAB. Позволяет работать с многомерными массивами данных, поддерживает основные математические методы, операции и структуры данных, такие как поддержка множества тензорных операций; работа с разреженными матрицами и поддержка ряда операций с ними; многочисленные методы линейной алгебры. Возможно использование архитектуры CUDA для мультипроцессорной работы с многомерными массивами данных.

Среда для глубинного обучения Caffe [5.2] написана на языке C++ и поддерживает интерфейс на языке Python. Эта среда поддерживает много типов машинного обучения, нацеленных в первую очередь на решение задач классификации и сегментации изображений. Caffe реализует сверточные нейронные сети, RCNN, долгую и краткосрочную память и полностью соединенные

нейронные сети. Для ускорения обучения может применяться система графических процессоров (GPU), поддерживаемая архитектурой CUDA.

Кроссплатформенная библиотека машинного обучения ML.NET [5.12] реализует такие методы машинного обучения, как классификация и регрессия. Платформа имеет свое API.

Решение Cognitive Toolkit (CNTK) [5.3] позволяет создавать сети глубокого обучения (DNN), сверточные сети (CNN), рекуррентные сети и сети с памятью (RNN, LSTM).

Выводы

Реализация предлагаемых подходов с целью дальнейшей оценки ФСО специалиста и прогнозирования его работоспособности предполагает создание на базе Стенда прототипа аппаратно-программного стенда для оценки ФСО специалиста и прогнозирования его работоспособности в различных условиях, в т.ч. в полевых и непосредственно на его рабочем месте. Авторами разработан вариант конфигурации такого Стенда. Его основной задачей является синхронная регистрация данных разной модальности о человеке с одновременной и также синхронной регистрацией данных о состоянии внешней среды. Требование синхронности обязательное. Его реализация предполагает синхронную регистрацию всех цифровых данных, видеоданных и акустических данных. Точность синхронизации должна быть не ниже 0,0001 с. На сегодня на рынке оборудования для синхронного многоканального сбора данных с указанными характеристиками присутствует небольшое число производителей относительно законченных модульных решений. Российским лидером является компания «Октава-ЭлектронДизайн» (Россия, [5.15]), мировым лидером является компания «National Instruments» (США, [5.21]).

Выполненная интеграция приборов в единый Стенд существенно расширяет возможности использования полиграфа «Диана-7М», видеокомплекса «Диана-Видео» и анализатора звуковых колебаний R&SUPV в исследовательских целях. Удобство снятия показателей с разных устройств позволяет оперативно работать с полученными данными, так как снятие происходит синхронно, вследствие чего не приходится тратить время на их сопоставление.

Прилагаемый сервер позволяет хранить большие массивы данных, благодаря этому полученный результат можно сразу направить на обработку и анализ. Способность стенда неинвазивно регистрировать показатели разной модальности расширяет возможности контроля и прогноза функционального состояния оператора. Дальнейшее развитие Стенда будет проводиться в направлении совершенствования аналитической обработки собранных данных на основе известных фреймов и платформ машинного обучения. Также планируется расширить функционал стенда в направлении синхронного предъявления стимульного материала аудиовизуальных модальностей.

Авторы допускают, что полученный ими набор речевых, визуальных и психофизиологических паттернов, характеризующих степень утомления будет в дальнейшем использован:

- в разработке методик диагностики утомления человека-оператора;
- в процессе медико-психологического сопровождения операторов сложных человеко-машинных систем;
- в разработке методик ранней диагностики пограничных функциональных реакций организма человека на основании его речевой деятельности;
- в дальнейшей перспективе развития Стенда предполагается поиск критериев дифференциальной диагностики голосового утомления, что позволит разработать мероприятия по профилактике голосового утомления у преподавателей, педагогов, вокалистов.

Благодарности

Авторы признательны инженерам-программистам и, одновременно, магистрантам Санкт-Петербургского государственного университета аэрокосмического приборостроения Антону Владимировичу Жеронкину, Артемию Алексеевичу Овсянникову и Виталию Александровичу Быстрову за их вклад в создание представленного в работе Стенда.

Библиографический список

5.1. *Baltrusaitis T.*, *Robinson P.*, *Morency L.-P.* OpenFace: An open source facial behavior analysis toolkit // 2016 IEEE Winter Conference on Applications of Computer Vision. — 2016. — P. 1–10.

5.2. Caffe: Convolutional Architecture for Fast Feature Embedding / Y. Jia [et al.] // CoRR. — 2014. — Vol. abs/1408.5093. — URL: http://arxiv.org/abs/1408.5093.

5.3. Cognitive Toolkit. — URL: http://www.microsoft.com/en-us/cognitive-toolkit/ (visited on 19.04.2019).

5.4. Detecting depression from facial actions and vocal prosody / J. F. Cohn [et al.] // 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops. — IEEE. 2009. — P. 1–7.

5.5. Dimensional information-theoretic measurement of facial emotion expressions in schizophrenia / J. Hamm [et al.] // Schizophrenia research and treatment. — 2014. — Vol. 2014. — P. 1–10.

5.6. *Ekman P.*, *Friesen W.*, *Hager J.* Facial action coding system. — Salt Lake City, 2002. — 527 p.

5.7. Facial expression in patients with bipolar disorder and schizophrenia in response to emotional stimuli: a partially shared cognitive and social deficit of the two disorders / G. Bersani [et al.] // Neuropsychiatric disease and treatment. — 2013. — Vol. 9. — P. 1137–1144.

5.8. Justification of the structure of an information system for analyzing the image of a human face for medical purposes / A. Yakovlev [et al.] // "Preventive Medicine 2017": Collection of scientific papers of the All-Russian Scientific and Practical Conference with international participation. — Publishing House of North-Western Medical University named. I.I. Mechnikov. 2017. — P. 306–312. — (In Russian).

5.9. *Keomany J.*, *Marcel S.* Active shape models using local binary patterns. — 2006. — 55 p.

5.10. *Kipyatkova I.*, *Ronzhin A.*, *Karpov A.* Automatic processing of colloquial russian speech. — Saint-Petersburg: GUAP, 2013. — 314 p. — (In Russian).

5.11. Median filtration. — URL: https://ru.bmstu.wiki/%D0%9C%D0%B5%D0%B4%D0%B8%D0%B0%D0%BD%D0%BD%D0%B0%D1%8F_%D1%84%D0%B8%D0%BB%D1%8C%D1%82%D1%80%D0%B0%D1%86%D0%B8%D1%8F (visited on 19.04.2019); (In Russian).

5.12. ML.NET. — URL: http://www.microsoft.com/net/apps/machinelearning-ai/ml-dotnet (visited on 19.04.2019).

5.13. *Naidenova X.*, *Ivanov V.*, *Yakovlev A.* Discretization of attributes with continious scales for extraction of concept knowledge from experimental data // Proceedings of 9th national conference on artificial intellect with foreign participation. — 2004. — Vol. 1. — P. 145–153. — (In Russian).

5.14. *Naidenova X.*, *Ivanov V.*, *Yakovlev A.* Discretization of Numerical Attributes and Extraction of Concept Knowledge from Data // Advances in Data Mining and Knowledge Discovery: Abstracts of Conference «Mathematical Methods for Learning». — 2004. — P. 54–56.

5.15. Octava ElectronDesign. — URL: http://www.octava.info/ (visited on 19.04.2019).

5.16. Octave. — URL: http://www.gnu.org/software/octave/ (visited on 19.04.2019).

5.17. *Omelchenko A.*, *A.V. F.* Evaluation of polynomial regression parameters based on set of realisations // Radioelectronics and informatics. — 2009. — Vol. 1. — P. 28–32. — (In Russian).

5.18. On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common / F. Weninger [et al.] // Frontiers in Psychology. — 2013. — Vol. 4. — P. 1–12. — DOI 10.3389/fpsyg.2013.00292.

5.19. Pletizmography // Big medical enciclopedy. Vol. 19 / comp. by B. Petrovsky. — M.: Soviet enciclopedy, 1974. — P. 194. — (In Russian).

5.20. *Rahmanenko I.*, *Mescheryakov R.* Analysis of indification attributes in speech data using GMM-UBM system of dictor verification // Proceedings of SPIIRAS. — 2017. — Vol. 3, no. 52. — P. 32–50. — (In Russian).

5.21. Recent Developments in OpenSMILE, the Munich Open-Source Multimedia Feature Extractor / F. Eyben [et al.] // Proceedings of the 21st ACM International Conference on Multimedia. — New York, NY, USA: Association for Computing Machinery, 2013. — P. 835–838. — URL: https://doi.org/10.1145/2502081.2502224.

5.22. Scikit-Learn. — URL: https://scikit-learn.org (visited on 19.04.2019).

5.23. Scikit-learn: Machine learning in Python / F. Pedregosa [et al.] // Journal of machine learning research. — 2011. — Vol. 12. — P. 2825–2830.

5.24. *Solodkov A.* Sport physiology: functional states of sportsmen and ways of their recovering. — Saint-Petersburg: National state university of physical culture, sport, health named after P.F. Lesgaft, 2015. — 94 p. — (In Russian).

5.25. Tensorflow. — URL: https://www.tensorflow.org/ (visited on 19.04.2019).

5.26. The experience of creating a hardware-software complex for multimodal registration of the functional state of a person / A. Yakovlev [et al.] // Modern contradictions and directions of development of aviation and space medicine. Materials of the scientific-practical conference "Actual problems of aviation and space medicine", dedicated to the 60th anniversary of the department of aviation and space medicine. — Military Medical Academy. 2018. — P. 196–201. — (In Russian).

5.27. The timbre toolbox: Extracting audio descriptors from musical signals / G. Peeters [et al.] // The Journal of the Acoustical Society of America. — 2011. — Vol. 130, no. 5. — P. 2902–2916.

5.28. Theano. — URL: www.deeplearning.net/software/theano/ (visited on 19.04.2019).

5.29. *Vachhani B. B.*, *Patil H. A.* Use of PLP cepstral features for phonetic segmentation // 2013 International Conference on Asian Language Processing. — IEEE. 2013. — P. 143–146.

5.30. *Velichko A.*, *Budkov V.*, *Karpov A.* Analytical review of computer paralinguistic systems for the automatic lie recognition in human speech // Informatic-Control Systems. — 2017. — Vol. 3, no. 90. — P. 30–41. — (In Russian).

5.31. *Yakovlev A.*, *Matytsin V.* Comprehensive diagnostics of vocal fatigue at the operators based on the use of machine learning techniques and computer vision // Preventive medicine-2018: a collection of scientific papers of the All-Russian scientific-practical conference with international participation. — Publishing House of North-Western Medical University named after I.I. Mechnikov. 2018. — P. 238–242. — (In Russian).

5.32. *Yakovlev A.*, *Matytsin V.* Generalized design of the study of audiovisual and psycho-physiological patterns that characterize the operator voice fatigue // Modern contradictions and directions of development of aviation and space medicine. Materials of the scientific-practical conference "Actual problems of aviation and space medicine", dedicated to the 60th anniversary of the department of aviation and space medicine. — Military Medical Academy. 2018. — P. 191–195. — (In Russian).

5.33. *Yakovlev A.*, *Naidenova X.* Concept of using technologies of big data in contemporary medicine // News of Russian Military-medical academy. — 2018. — Vol. 37, no. 1. — P. 17–23. — (In Russian).

5.34. *Yakovlev A.*, *Naidenova X.* Ontology as knowlege systematization instrument about evaluation devices for monitoring of servicemen functional state of organism // Actual problems of protection and security: Proceedings of XIX All-russian scientific-practice conference RARAN (4-7 april 2016). — 2016. — Vol. 7, no. 1. — P. 400–403. — (In Russian).