

РАЗДЕЛ IV. АКТУАЛЬНЫЕ НАПРАВЛЕНИЯ ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ ТРУДА, ОРГАНИЗАЦИИ И УПРАВЛЕНИЯ В ТРЕТЬЕМ ДЕСЯТИЛЕТИИ XXI ВЕКА

ГЛАВА 19. МЕТОД ПРЯМОЙ ОЦЕНКИ ЛИЦЕВЫХ ЭКСПРЕССИЙ НА ВИДЕОЗАПИСИ: ОТ ВОСПРИЯТИЯ ЭКСПЕРТА К КОМПЬЮТЕРНОМУ ЗРЕНИЮ

Постановка проблемы

С развитием компьютерных технологий и применением новейших методов обработки данных в психологии труда, психодиагностике, HR-бизнесе, нейромаркетинге, судебно-психологической экспертизе и других областях практической психологии все более широкое распространение получает использование высококачественных инструментов видеорегистрации поведения человека и автоматического анализа этих данных для оценивания человека по его экспрессивным реакциям. Например, попытки автоматизированного анализа видеointервью кандидатов на занятие вакантных должностей при массовом рекрутинге персонала в крупных российских компаниях становятся все более и более очевидными (см., например, платформу www.cvc.ru). После общения с коллегами и публикаций в прессе нам известны некоторые методические наработки таких крупных компаний, как Сбербанк, Северсталь. Насколько эти разработки «жизнеспособны» и применимы на практике, вопрос особый, поскольку научно обоснованных и эмпирически подтвержденных свидетельств эффективности своих программных продуктов компании-разработчики, как правило, не предоставляют, а доступные демоверсии вызывают у психологов-профессионалов улыбку. Как правило, разработчики лишь пишут о том, что они использовали нейросетевые технологии.

Желание руководства ряда IT-компаний и наших коллег-психологов разрабатывать новые методы оценки психологических особенностей человека, используя не только и даже не столько традиционные самооценочные методы, но и более *объективные* и *надежные* критерии оценивания человека вполне понятны, поскольку, например, традиционные опросные методы страдают таким важным недостатком, как влияние социальной желательности в ответах респондентов на результаты психодиагностических оценок.

Одним из направлений компьютерной психодиагностики поведения человека является *анализ лицевых экспрессий*, зарегистрированных на видеозаписи, например, в ходе дистанционного кадрового интервью. Хорошо известно, что по изменениям на лице человека можно не только определять выраженность эмоций, но и оценивать эмоциональную напряженность, стресс, неуверенность в ответах или даже попытку исказить или скрыть предоставляемую информацию (Ekman, Friesen, Hager, 2002a; Rosenberg, Ekman, 2020; Davidson, Scherer, Goldsmith, 2003; Calvo et al., 2015; Davis,

Panksepp, 2018; Баев, Гусев, 2020). Особо подчеркнем, что перспективность использования результатов профессионального анализа лицевых экспрессий пока резко контрастирует с неготовностью нашего профессионального сообщества это делать, т. е. качественно и объективно оценивать видео материал и выносить на основании проведенного анализа обоснованные заключения.

Укажем на две важные причины такого положения дел:

1. Крайне высокая трудоемкость оценки мимической активности (МА) даже опытным экспертом. Так, по данным наших коллег, имеющих сертификат о прохождении теста на владение FACS в Paul Ekman Group и более или менее постоянно занимающихся экспертизой МА по видеозаписям, общая трудоемкость эксперта в самом лучшем случае оценивается как 1 : 10 по отношению к длительности анализируемой видеозаписи, а весьма часто доходит до пропорции 1 : 30. То есть работа квалифицированного эксперта очень трудоемка.

2. Вторая проблема – это отсутствие (по крайней мере, в нашей стране) необходимого числа опытных экспертов. По имеющейся у одного из соавторов настоящей публикации информации, известно лишь о трех сертифицированных на международном уровне специалистах. Не так уж много высококвалифицированных специалистов и в других странах. Например, наш опыт работы с базой данных (БД) лицевых экспрессий SAMM (Chuin et al., 2020), созданной коллегами из Манчестера, показал не очень высокий уровень работы экспертов, привлекавшихся для анализа включенных в эту БД видеоматериалов, особенно регистраций микровыражений лица.

Таким образом, особую важность приобретает разработка программного обеспечения (ПО), которое позволит анализировать МА с высокой точностью, сравнимой с результатами работы опытного эксперта. Нельзя сказать, что на рынке ПО нет такого инструментария, например, среди психологов неплохо себя зарекомендовали разработки зарубежных и даже отечественных (Нейроботикс) компаний. Один из авторов давно работает с ПО Face Reader голландской компании Noldus (версии от 4.0 до 8.2), другой из нас хорошо знаком с ПО iMotions. Тем не менее укажем на то, что использование этого ПО создает опасность получения некачественных оценок не только отдельных мимических проявлений в виде двигательных единиц или Action Units (AUs) по системе FACS (Ekman, Friesen, Hager, 2002b), но даже базовых эмоций. Иногда искажения в результатах компьютерного анализа бывают велики и не соответствуют оценкам экспертов, основанным на строгих критериях.

Таким образом, цель публикации результатов научной и практической работы нашей группы состоит в том, чтобы поделиться с коллегами достижениями в области разработки методологии и технологии автоматического оценивания МА с помощью созданного нами ПО с

акцентом на его использование в научных исследованиях и при решении практических задач оценки поведения человека.

Как можно анализировать лицо?

Разработанная нами методология оценки МА основана на известной системе FACS (Facial Action Coding System), разработанной П. Экманом и его коллегами (Ekman, Friesen, Hager, 2002a, б) и получившей самое широкое распространение как способ универсального описания движений, наблюдаемых экспертом на поверхности лица. Единицей анализа в системе FACS является двигательная единица – AU, как характерное изменение поверхности лица в форме отдельного движения (например, губ, бровей, век, носогубных складок). Согласно руководству по FACS, в ходе визуального анализа лица человека выделяют ограниченный набор AUs (рис. 19.1). Определенная композиция этих базовых единиц анализа МА позволяет выделить семь базовых эмоций. Особо отметим, что система FACS не задает теоретико-методологические рамки для оценки мимических выражений эмоций, тем более выделения определенного набора базовых эмоций. В современной литературе есть известная дискуссия о том, какие эмоции отражаются в МА, какие из них базовые и сколько их (см., например: Rosenberg, Ekman, 2020; Davidson, Scherer, Goldsmith, 2003; Davis, Panksepp, 2018; Фельдман Барретт, 2018). Поэтому мы акцентируем внимание читателей на том, что FACS – это система *инвариантного* описания МА, инструмент, который можно использовать для выделения на лице выражения эмоций и других сложных паттернов МА вне зависимости от того, как разные авторы понимают и операционализируют задачу обнаружения эмоций на лице человека. Система FACS *не связана* с какими-либо теоретическими предположениями о природе выражения лица, она основана на очевидной и эмпирически подтвержденной связи активности отдельных лицевых мышц и мышечных с групп с соответствующими изменениями поверхности лица. Это важно, поскольку даже профессиональные психологи нередко смешивают эти две проблемы – описание МА и ее категоризацию в виде эмоциональных экспрессий. FACS – это не более чем инструмент для базовой оценки МА как способ систематической фиксации в поле восприятия и категоризации определенных движений поверхности лица. Например, AU2 – это поднятие внешней части брови, AU41 – опускание надпереносья, а AU20 – растяжение губ (см. номера отдельных AUs на рис. 19.1). Из сочетаний AUs в соответствии с рекомендациями создателей FACS выделяют базовые эмоции, например: 1) радость определяется как сочетание AUs 6+12; 2) грусть – как сочетание AUs 1+4+11+15 или 1+4+15 (Ekman, Friesen, Hager, 2002a).

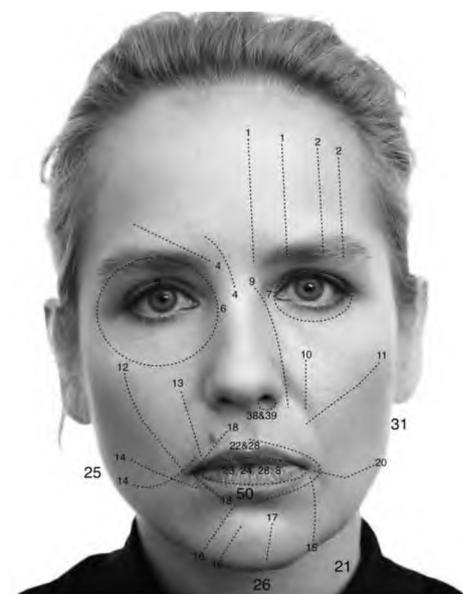


Рис. 19.1. Основные двигательные единицы, выделяемые на лице человека

За выделение AUs всегда лежит активность лицевых мышц (рис. 19.2). Поэтому за обнаружением экспертом появления или исчезновения той или иной AUs стоит напряжение или расслабление одной или нескольких мышц, активность которых приводит к наблюдаемым экспертом смещениям поверхности лица. Например, за AU2 стоит активность латеральной части лобной мышцы, за AU41 – сокращение мышцы гордецов, а за AU20 – сокращение так называемой мышцы смеха.

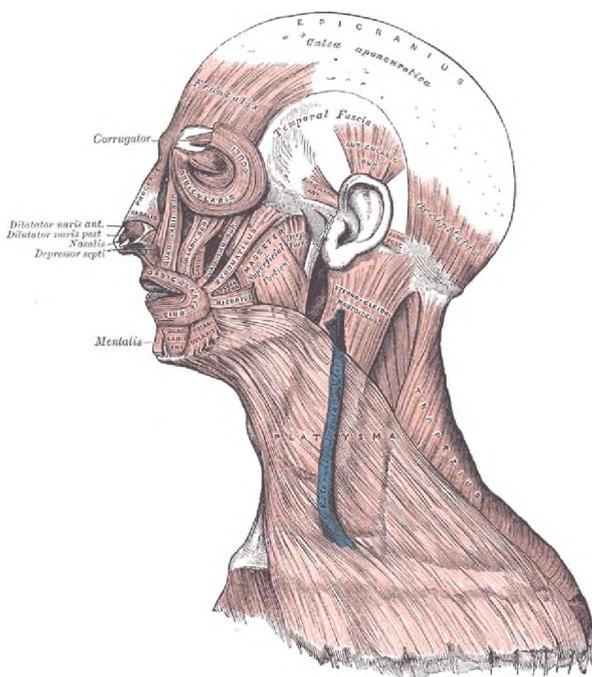


Рис. 19.2. Отдельные лицевые мышцы, активность которых приводит к появлению на лице двигательных единиц
(по: https://en.wikipedia.org/wiki/Facial_muscles#/media/File:Gray378.png)

Два подхода к автоматическому анализу МА на лице человека по видеозаписи

Автоматизированный анализ МА по видеозаписи или в реальном времени представляет собой особую проблему. Суть ее заключается в том, что компьютерная система должна с приемлемой точностью и надежностью выделять разные проявления МА, например, отдельные AUs и базовые эмоции. В литературе описаны два основных подхода к решению этой задачи. Рассмотрим их подробно.

Так называемый *селективный подход* основан на поиске соответствия определенных выражений лица образцам (как вариантов МА) в ограниченном наборе базовых эмоций или AUs. Его характерные особенности и проблемы состоят в следующем:

1. Используются нейросети, обученные на ограниченной и, как правило, маловариативной выборке фотографий лиц людей с выражением шести-семи базовых эмоций и отдельных AUs. Большая проблема состоит в том, для обучения нейросети нужны обучающие выборки изображений лицевых экспрессий, аннотированные опытными экспертами, а получить их крайне сложно. Тем более создать такие выборки, где были в большом количестве представлены разные AUs, микровыражения лица, не говоря уже о том, что эти данные нужно собрать на людях разных этнических групп.

Серьезным недостатком этого подхода является то, что эксперты, аннотирующие фотографии, составляющие обучающую выборку, оценивают каждое фото на предмет выраженности в нем эмоции или какой-либо AU *статически*, по впечатлению, а не анализируют видеозапись, оценивая мимические проявления *в динамике*. На наш взгляд, это серьезная ошибка, поскольку AUs – это двигательные феномены, согласно FACS, у них есть начало, максимум выраженности и конец. Нам пока не известны разработки, в которых нейросети обучались на видеозаписях и их оценках экспертами. По крайней мере, ПО Face Reader и iMotions разрабатывались с использованием большой базы данных фотографий.

2. В силу особенностей используемых компьютерных алгоритмов и ограниченности обучающих выборок быстрые движения (микровыражения) лица не определяются или определяются плохо.

3. В рамках данного подхода у 7–10 % людей в силу их индивидуальных особенностей и/или неврологических проблем невозможно оценить эмоции по выражению лица, поскольку эмоции не выражены. В силу невысокой точности автоматического обнаружения AUs как базовых единиц МА и ограниченности их набора у таких людей описание МА весьма проблематично.

4. Необходим учет расовой принадлежности лица анализируемого человека, что требует использования для обучения нейросети выборки изображений лиц разных этносов. В настоящее время обучающие выборки созданы для очень ограниченного числа рас (см., например, возможности ПО

Noldus Face Reader, где предлагается выбор двух рас – кавказская и Юго-Восточная Азия).

5. Как правило, необходимость оценки индивидуальных особенностей МА разных категорий респондентов при решении разных исследовательских или практических задач (например, нейромаркетинг или анализ видеointервью при кадровом отборе) требуют ручной настройки точности срабатывания используемых нейросетевых алгоритмов. Например, ПО компании Affectiva предлагает пользователям использовать для решения разных задач до трех уровней срабатывания алгоритма обнаружения эмоций на видеозаписи.

6. Проблематичность описания и категоризации МА детей до 5–7 лет.

Комплексный подход основывается на выделении и анализе сочетаний AUs во времени. Его особенности следующие:

1. Принципиальное использование международной системы FACS и непосредственное выделение AUs как основных единиц анализа МА позволяют описывать все возможные выражения лица.

2. Возможность оценить быстрые движения лица, асимметрию МА, в том числе микровыражения эмоций. Эта возможность обеспечивается выделением базовых единиц МА – AUs, а на их основе – самых разных мимических феноменов. Она ограничена лишь качеством видеозаписи и точностью работы компьютерного алгоритма анализа МА.

3. Возможность оценки индивидуальной вариабельности МА, например, обнаружение мимических гиперкинезов, маньеризмов, сложных паттернов МА. Эти мимические феномены оцениваются как появление соответствующих движений на поверхности лица, т. е. как отдельные AUs или их комбинация, а не как итоговое «впечатление» нейросетевого классификатора.

4. Независимость от расовой принадлежности человека, поскольку данный подход не ограничен особенностями обучающей выборки фото-, видеоизображений лиц.

5. Не требуется индивидуальная настройка компьютерной системы в соответствии с особенностями анализируемой выборки и решаемой задачи, поскольку в ходе анализа конкретной видеозаписи оценивается индивидуальная динамика МА конкретного человека. Эксперт или компьютерная система принимают решение с опорой на текущие изменения в МА, основываясь на впечатлении об интенсивности и характере видимых изменений или заданных пороговых значениях этих изменений в рамках всей видеозаписи или ее части.

6. Нет возрастных ограничений по оценке базовых проявлений МА, поскольку описываются появление AUs, а не эмоции.

Разработанный нами способ компьютерного анализа МА базируется на последнем подходе.

При создании конкретных алгоритмов автоматизированного анализа МА мы использовали ряд важных методических принципов, дающих нам очевидные преимущества:

1. *Прямая оценка* перемещений поверхности лица для обнаружения AUs и принципиальный отказ от использования нейросетей для классификации МА как метода их косвенного оценивания.

2. *Моделирование восприятия* экспертом особенностей перемещений поверхности лица при обнаружении отдельных AUs с учетом топографии движений лицевой поверхности, характерной для реального движения асимметрии, распределения пространственного внимания в ходе визуального наблюдения за движениями лица другого человека, характера процесса принятия решения о наличии/отсутствии AUs в соответствии с рекомендациями FACS.

Особо отметим, что при создании ПО были разработаны пять базовых *оригинальных* (авторских) процедур компьютерного зрения (CV), специфически ориентированных на анализ движений поверхности лица в отдельных его частях и характерные особенности проявления разных AUs. Необходимость создания новых процедур CV была вызвана тем, что стандартные процедуры Open CV (открытые для использования процедуры компьютерного зрения) представляли собой универсальные способы преобразования изменений светового потока в паттерны движения и не были ориентированы на анализ характерных сдвигов поверхности лица. Мы исходили из того, каким образом изменения света, попадающие в зрительную систему наблюдателя, преобразуются в двигательные перцептивные инварианты (Гибсон, 1988). Фактически вновь разработанные процедуры CV (facelines, regions, texture, eyebrowtracking, eyebrowform) – это поиск максимального соответствия механизмов зрительного восприятия (как сложного перцептивного действия) современным способам анализа изменений светового потока, зарегистрированного видеокамерой. Для более тонкого анализа движения глаз и движения губ используется ряд дополнительных процедур компьютерного зрения.

Не вдаваясь в технические подробности, отметим, что эти новые процедуры по-разному оценивают изменения света в разных частях лица, разделенного нами на 14 симметричных зон (сегментов) анализа (рис. 19.3), а также в отдельных мини-сегментах с учетом вероятности появления AUs на местах перекрытия этих зон. Фактически были созданы компьютерные процедуры, анализирующие сокращение и растяжение поверхности лица в тех направлениях, которые соответствуют каждой из AUs. Например, понятно, что метрика для обнаружения движений внутренних частей бровей вверх и друг к другу (AUs 1+4) – это совсем другая метрика по сравнению с горизонтальным сжатием нижних век (AU 7).

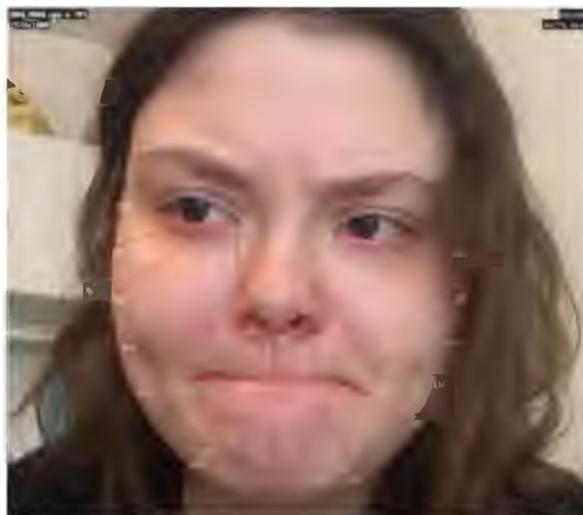


Рис. 19.3. Базовые сегменты анализа поверхности лица, используемые ПО F2F Emotion Studio

Фактически при выделении отдельных AUs нами был реализован принцип многослойного анализа МА:

- разделение видеозаписи на отдельные кадры, выделение артефактов, построение сегментной карты лица по автоматически выделенным 68 реперным точкам, соответствующим анатомии лица;

- анализ изменений распределения света от кадра к кадру в разных сегментах лица, выделение значимых изменений, интерпретация этих изменений на поверхности лица в виде сдвигов кожи, складок, выпуклостей и вогнутостей; этот нижний уровень представляет собой сложный 3-х этапный анализ изменений поверхности лица, результатом которого является первичное выделение значимых мимических событий на фоне общего «шума».

- сопоставление выделенных изменений друг с другом в соответствии с анатомическим строением лица и схемой FACS;

- выделение начала и окончания отдельных AUs;

- выделение эмоций и других сложных мимических паттернов;

- расчет интегральных показателей МА, например: частота AUs в единицу времени, валентность, асимметрия, сочетание AUs и речи.

- визуализация результатов анализа на временной оси и построение таблицы результатов анализа МА;

- сложная категоризация паттернов МА как характерных проявлений мимического поведения: истинные эмоции, эмблемы эмоций, придание эмоциональной модальности речевому высказыванию, индивидуальные мимические привычки, выделение значимых фрагментов речи и др.

В ходе нашей работы была создана система правил, на основании которых «сырые» данные об изменении света преобразуются в движение поверхностей, а они, в свою очередь, – в AUs. Фактически эти правила и их соотношения могут быть прообразом нового компьютерного языка для анализа МА.

Эмпирическая апробация разработанной технологии анализа МА

В настоящее время разработано ПО для проведения полностью автоматизированного анализа МА по видеозаписям с разрешением не хуже HD, на которых лицо человека записано анфас (с небольшими поворотами головы влево-вправо, вверх-вниз) и занимающее не менее 15–20 % площади кадра.

Для исследований и дальнейшей разработки компьютерных технологий анализа МА создана компьютерная система *F2F Emotion Studio*, работающая в среде Windows. Она позволяет не только обрабатывать видеозаписи, но и тонко настраивать правила обнаружения отдельных AUs, выделения базовых эмоций и других показателей МА (в том числе сложных паттернов МА). Всего с высокой точностью обнаруживаются 18 основных двигательных единиц: AU 1, 2, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 24, 28, 41, 44, 43, 45, а также семь базовых эмоций: радость, удивление, презрение, печаль, отвращение, страх, гнев. Для решения практических задач разработано прикладное ПО – *ЭмоРадар*, работающее под Windows и Linux (включая защищенную ОС Astra Linux). Все варианты ПО имеют возможность работать с мощным удаленным сервером, что значительно повышает скорость обработки видеозаписей. ПО *ЭмоРадар* позволяет обрабатывать видеозаписи и просматривать на временной оси изменения МА, отображая на экране монитора разные эмоции и AUs.

Для эмпирической верификации точности и надежности работы созданного ПО мы использовали две базы данных микро- и макровыражений лица – CASME II (Yan et al., 2014) и SAMM (Yap et al., 2020).

Из каждой базы данных было случайным образом взято и проанализировано по 30 видеозаписей. Более 90 % AUs было обнаружено с вариациями оценок времени их начала и окончания от 12 до 25 % по сравнению с оценками экспертов, привлеченных к работе с этими базами данных. Оставшиеся 9–10 % записей, на наш взгляд, были не всегда корректно закодированы экспертами в соответствии с критериями FACS. Это было особенно типично для базы данных SAMM, где мы обнаружили очевидные ошибки кодирования. После внесения необходимых, по нашему мнению, коррекций в их оценки, точность обнаружения AUs и базовых эмоций при работы нашей компьютерной системы достигла 97 % (Gusev, Baev, Kremlev, 2021).

Кроме того, было проанализировано множество видеозаписей с выступлениями государственных и политических деятелей разных стран, которые были предварительно вручную аннотированы по системе FACS М.С. Баевым как сертифицированным экспертом по анализу лицевых экспрессий. С учетом выявления маньеризмов, индивидуальных мимических привычек, а также мимических феноменов эмоционального подчеркивания речи наше ПО также выявляло не менее 95 % AUs и базовых эмоций.

На рис. 19.4 представлены результаты анализа *макровыражений* лица одной из наших испытуемых. Автоматически обнаружен ряд AUs и две

базовых эмоции – презрение (Contempt) и отвращение (Disgust). На рис. 19.5 показан результат анализа одной из видеозаписей, демонстрирующей обнаружение *микровыражения* лица. Автоматически была обнаружена асимметричная AU20 в левой половине лица длительностью 180 мс и, соответственно, микровыражение эмоции страха. Это очень быстрое мимическое событие, которое не без труда замечает не каждый эксперт.

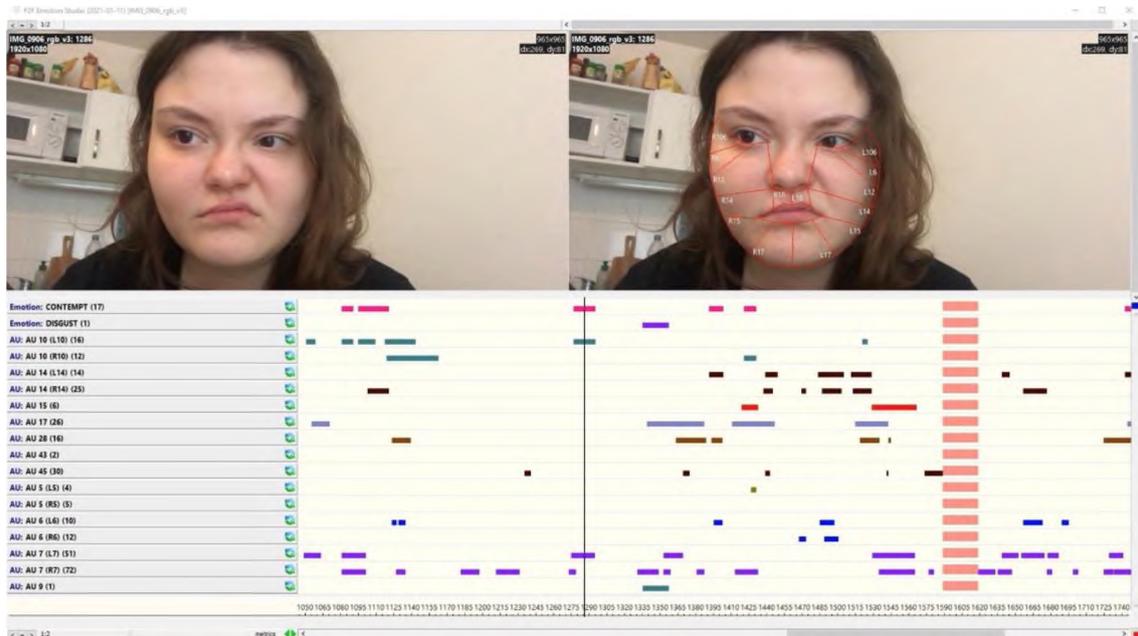


Рис. 19.4. Результат компьютерного анализа МА испытуемой М.Б. с помощью ПО F2F Emotion Studio. Видеозапись взята из базы данных Лаборатории исследования поведения компании «Лицом к лицу», испытуемая М.Б. (запись IMG_0906_rgb_v3, fps = 30)



Рис. 19.5. Результат компьютерного анализа МА испытуемого Sub10 с помощью ПО F2F Emotion Studio. Видеозапись взята из базы данных CASME II (запись sub10_EP13_01, fps = 200)

Практическое использование технологии анализа МА в психологии и смежных областях

Разработанная нами технология анализа МА, основанная на FACS как универсальной методологии описания МА, позволяет использовать ее в рамках различных схем интерпретации, поскольку обеспечивает всеобъемлющее и инвариантное *описание* мимики как таковой. В зависимости от схемы интерпретации результатов оценки МА и стоящей перед психологом-исследователем или психологом-практиком цели возможно решать разные задачи:

1) оценка изменения эмоционального состояния человека при воздействии на него различных стимулов; оценка и подбор стимульного материала в виде записи лицевых экспрессий человека; нейромаркетинг и исследование потребительского поведения;

2) оценка изменения эмоционального состояния человека в ходе выполнения профессиональной деятельности; анализ бдительности при выполнении сложных задач, требующих контроля функционального состояния оператора;

3) диагностика неблагоприятных психических состояний, например, депрессий (Ellgring, 2008);

4) анализ особенностей коммуникации, включая ситуации речевого общения; оценка эмоционального выражения лица человека на зрителя (собеседника), в том числе в ходе публичных выступлений и через средства массовой информации;

5) судебно-психологическая экспертиза видеозаписей; диагностика скрываемого аффекта, попытки искажения или скрытия информации (Баев, Гусев, 2020);

6) дистанционная автоматизированная оценка видеоинтервью при приеме на работу;

7) получение данных для обучения нейросетей, ориентированных на оценку мимики отдельного человека.