

Для цитирования: Гусев А. Н., Енгальчев В. Ф., Захарова Н. А. Компьютерные технологии оценки голоса и лицевых экспрессий в анализе аудио- и видеоматериалов // *Armenian Journal of Mental Health*. — 2018. — Т. 9, № 1 /Supplement. — С. 70–73.

А.Н. Гусев, В.Ф. Енгальчев, Н.А. Захарова

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ОЦЕНКИ ГОЛОСА И ЛИЦЕВЫХ ЭКСПРЕССИЙ В АНАЛИЗЕ АУДИО- И ВИДЕОМАТЕРИАЛОВ

Ключевые слова: психологическая экспертиза аудио- и видеозаписей, двигательное и вербальное поведение, эмоциональное выражение лица, вокальные параметры голоса.

При проведении психологической экспертизы аудио- и видео материалов целесообразно учитывать преимущества использования компьютерных средств оценивания поведения человека, особенно тех инструментов, которые реализуют бесконтактные методы диагностики психоэмоционального состояния человека, достоверности/недостоверности предоставляемой им информации.

Психологический анализ аудио- и видеозаписей прежде всего реализуется методом наблюдения. Из этого следует, что, с одной стороны, наблюдение – это процедура сбора непосредственно воспринимаемых психологом данных, а, с другой стороны, выбор способа их фиксации и схемы интерпретации наблюдаемых событий в соответствии с целью экспертного исследования. К наиболее часто встречающимся трудностям, в целом снижающим надежность экспертных заключения и их обоснованность, можно отнести то, что метод наблюдения, во-первых, является довольно трудоемким, во-вторых, традиционно исследователь одновременно наблюдает за поведением, записывает свои наблюдения и отмечает время наступивших событий (что, естественно, влияет на качество наблюдения), и, в-третьих, как правило, нет интеграции с другими методами анализа поведенческих реакций человека. Очевидно, что последнее не позволяет применять комплексный подход к анализу различных поведенческих реакций исследуемого лица.

Последние годы все большую популярность среди психологов приобретают компьютерные системы, разработанные голландской компанией Noldus Information Technology (www.noldus.com, см. публикации на сайте компании). Специфика реализации метода наблюдения с помощью системы **The Observer XT** как удобной среды, интегрирующей разнообразные показатели активности человека при проведении экспертизы видеоматериалов описана нами ранее [2], и ниже мы не будем на этом останавливаться подробно. Отметим лишь, что она позволяет эксперту-психологу автоматизировать и формализовать ряд важных процедурных моментов, что способствует повышению валидности

и надежности данного метода по сравнению с традиционными способами наблюдения. Это обеспечивается за счет последовательной реализации в рамках одной компьютерной системы всех этапов наблюдения, начиная от разработки дизайна исследования до количественного анализа данных и/или визуализации полученных результатов.

Компьютерная система **FaceReader**, также разработанная компанией Noldus Information Technology совместно с компанией Vicar Vision, предназначена для выявления и анализа интенсивности лицевых экспрессий человека. FaceReader автоматически распознает на лице по видео записи или фотографии выражение шести базовых эмоций: счастье, грусть, гнев, удивление, испуг, отвращение и нейтральное эмоциональное состояние. Алгоритмы работы данной компьютерной системы основаны на результатах многолетних фундаментальных и прикладных исследований известного американского психолога Пола Экмана [3]. Кроме этого, данное программное обеспечение способно оценивать изменения мимики, фиксируя состояние глаз, рта и бровей: открытый или закрытый правый/левый глаз, открытый или закрытый рот, поднятые, опущенные брови или нейтральные брови; также отслеживается общее направление взгляда испытуемого и положение головы. В новейшей версии FaceReader`а появилась возможность оценивать по изменению освещенности лица активность нескольких мышечных групп (Action Units) и частоту сердечных сокращений. Эти возможности, на наш взгляд, существенно расширяют диагностические возможности данного инструмента.

FaceReader использует современные алгоритмы оценки выражения лица, включая Deep neuronal network. Это сложная задача, поскольку анализ видео записей выражений лиц разных людей, сделанных в конкретных условиях освещения, сталкивается с проблемой большой вариабельности в положении головы человека относительно камеры, а также освещения лица. В работе данного программного обеспечения последовательно реализуются три основных алгоритма, обеспечивающих классификацию поступающей информации при анализе видеозаписей: *обнаружение лица на анализируемом изображении, построение модели лица на основе расположения на лице человека почти 500 ключевых точек, классификация выражений лиц* при помощи алгоритма обучения искусственной нейронной сети на основе выборки, состоящей из более 10000 проанализированных ранее фотографий лиц различных людей. По сведениям, полученным от разработчиков этой системы, в последующие версии этого программного обеспечения будет включен модуль анализа паттернов активности разных двигательных единиц. В этом направлении также работает и наша исследовательская группа.

Подчеркнем, что сравнение работы FaceReader`а с оценками специально обученных экспертов (так называемые «фейс-кодеры») показало, что данный инструмент может определять степень выраженности базовых эмоций с точностью в среднем 95%, а по ряду

эмоций превышает 97%. На наш взгляд эта функция весьма уникальна и полезна в виду того, что опытных специалистов в мире мало, и их привлечение к проведению экспертизы видео записей крайне ограничено.

Для повышения надежности экспертизы видеозаписей крайне важно, что имеется возможность простого экспорта результатов анализа изменения выражения эмоций и других лицевых экспрессий исследуемого лица, выполненного с помощью системы FaceReader, в компьютерную систему The Observer XT. Это позволяет проводить оценку поведения человека по видеозаписи, одновременно наблюдая детальную динамику изменения знака и интенсивности эмоций и мимики, включая детекцию интенсивности выраженности Action Units в соответствии с международной системой кодирования эмоций FACS [3].

Далее рассмотрим преимущества использования мало известной экспертам-психологам компьютерной системы LVA-i (Layered Voice Analysis, многоуровневый анализ голоса), разработанной компанией Nemesysco (Израиль) для оценки тонких изменений в голосе исследуемого лица на предмет обнаружений признаков неискренности и различных проявлений эмоционального напряжения и когнитивного стресса. Данное программное обеспечение позволяет выделять в аудио записи или аудио канале изучаемой экспертом видео записи целый ряд информативных акустических показателей, несущих информацию об уровне неискренности исследуемого лица, напряженности когнитивных процессов, эмоциональном возбуждении [4, 5]. Суть этого метода состоит в том, что, если в голосе респондента повторяются некоторые показатели эмоционального или когнитивного стресса, то это свидетельствует о *риске*. Риске того, что говорящий испытывает дискомфорт, его функциональное состояние отличается от нормы (базовой линии), он взволнован, огорчен, затрудняется ответить на вопрос, переживает по этому поводу, пытается скрыть какую-то эмоциогенную информацию. Результаты анализа голоса не зависят от содержания речи респондента и языка, на котором он говорит, но они отражают его эмоциональное состояние во время произнесения речевого сообщения с учетом тонкой динамики индивидуальных особенностей его голоса – темпа, громкости, тембра.

Алгоритмы, используемые данной системой, значительно отличаются от традиционного фонетического анализа голоса, они основаны на анализе произвольно неконтролируемых изменений вокальных характеристик, происходящих в интервалах времени менее 10 мс. В ходе обработки речевого сигнала происходит выделение ряда последовательных фрагментов (0,3 ... 2 с), соответствующих периодам изменения амплитудно-частотных характеристик звукового сигнала (151 базовый параметр). В ходе многолетних эмпирических исследований автор-разработчик этой технологии – А. Либерман, собрал нормативные данные, характеризующие степень проявление в голосе различных

изменений психического состояния человека: уровень общего возбуждения, эмоциональное напряжение, разные проявления стресса, напряженность когнитивных процессов, концентрация, антиципация, смущение и др. (см. материалы сайта www.nemesysco.com). На основе сравнения целого ряда показателей система LVA-i оценивает интегральные показатели изменений, происходящих в голосе обследуемого лица, например, показатель SOS – сказать или промолчать. Также оценивается несколько обобщенных индексов, характеризующих проявления в голосе различных типов утаивания информации – от активного противодействия эксперту до пассивных ответов уклончивого характера.

Система LVA-i оценивает интегральный показатель общего риска сообщаемой информации, что позволяет использовать ее при проведении экспресс-анализа звучащей речи исследуемого лица. В результате проведенной офлайн обработки звукового файла данных эксперт получает ранговые оценки - высокий, средний или низкий уровни, выраженности в голосе испытуемого нескольких показателей изменения когнитивных процессов (внимание, обдумывание информации, уровень антиципации) и проявлений стресса. В зависимости от специфики задаваемых экспертом вопросов данная компьютерная система позволяет оценить по изменениям в голосе уровень искренности, лояльности, употребление алкоголя и наркотиков, склонность к азартным играм и др.

Качественные и количественные данные анализа голоса могут быть также импортированы в качестве внешних данных в систему The Observer XT и использованы в комплексном анализе поведения обследуемого лица наряду с изменениями его движений и эмоций (см. табл. 1).

Результаты проведенных нами лабораторных экспериментов и полевых исследований, представленных в докладе как отдельные кейсы, показали продуктивность использования комплекса указанных показателей в качестве надежных индикаторов изменения поведения при анализе видеозаписей бесед в ситуациях оценки его возможной причастности к краже денег, передаче третьим лицам строго конфиденциальной информации, употреблению алкоголя и наркотиков, недобросовестного исполнения служебных обязанностей.

Как было отмечено выше, очень важно проводить совместный анализ эмоциональных проявлений исследуемого лица, одновременно используя информативные показатели лицевой экспрессии и голоса. В качестве иллюстрации в таблице 1 представлены результаты анализа появления в голосе Ю. Скрипаль показателей эмоционального и когнитивного стресса (обобщенные индексы High Anticipation, Stress, High Risk) и характерной активности двух двигательных единиц на ее лице – AU 01 (Inner Brow Raiser) и AU 04 (Brow Lowerer) во время ее первого официального видео интервью. Проведенный анализ соответствует отрезку времени с 68.5 с до 83.0 с, когда она произносила указанные в таблице слова. Данные,

представленные в таблице могут быть легко импортированы в систему The Observer XT и представлены на одной временной оси вместе с анализируемой экспертом видео записью.

Табл. 1. Соотношение во времени параметров анализа голоса и двигательных единиц лица в видео обращении Ю. Скрипаль.

Показатели голоса и двигательные единицы	В дальнейшем я надеюсь вернуться домой в мою страну	Я решила прервать свою реабилитацию и сделать это короткое заявление, чтобы прояснить несколько вопросов.	Я убедительно прошу уважать мою личную жизнь и личную жизнь моего отца
High Anticipation	71.60-72.80.		
Stress	70.00-71.00.	77.80-80.00.	80.00-81.80.
High Risk			81.80-82.20.
AU 01 A	71.70-71.80.	76.80-76.95,77.25-77.60.	78.50-78.55,78.60-78.80, 79.30-79.70, 79.90-80.00, 80.10-80.40, 81.40-81.60.
AU 01 B	70.30-70.40, 71.50-71.70.	71.80-72.00,72.80-72.90,76.20-76.55, 76.60-76.80.	78.80-79.30,80.40-80.70,81.20-81.40, 81.60-81.85.
AU 01 C	68.50-70.30, 70.40-71.50.	72.00-72.80,72.90-75.70,75.90-76.20,76.55-76.60	80.70-81.20, 81.85-83.00.
AU 01 D		75.70-75.90.	
AU 04 A	68.53-69.20, 69.75-70.10, 70.80-70.90.	74.10-74.20,74.30-75.10,75.60-76.10.	82.00-82.20.
AU 04 B	68.50-68.53, 69.20-69.75.	75.10-75.60.	82.20-82.0.
AU 04 C			85.50-83.70.

Примечание: индексы А, В, С, D означают интенсивность выраженности соответствующих двигательных единиц.

Литература

1. *Багмет А.М.* и др. Методика выявления психологических признаков достоверности/недостоверности информации, сообщаемой участниками уголовного судопроизводства (по видеоматериалам следственных действий и оперативно-разыскных мероприятий). М., 2017.

2. *Гусев А. Н., Енгальчев В. Ф., Захарова Н. А.* Современные тренды в использовании программно-аппаратных средств при оценке психоэмоционального состояния человека // Аппаратные средства в психологической подготовке / под ред. А.Г. Караяни, С.И. Данилова. — Военный университет; Школа современных психотехнологий Москва, 2018. — С. 110–117.

3. Эрман П. Психология лжи / Пер. с англ. Н. Исуповой, Н. Мальгиной, Н. Миронова, О. Тереховой. — СПб. Питер, 2010. — 270 с.

4. Elkins A.C., Burgoon J., Nunamaker J. Vocal Analysis Software for Security Screening: Validity and Deception Detection Potential / *Homeland Security Affairs*, DHS Centers of Excellence Science and Technology Student Papers (March 2012).

<https://www.hsaj.org/articles/213>

5. Manchireddy B.; Sadaf S., Kamalesh J. Layered Voice Analysis Based Determination of Personality Traits / *Australasian Medical Journal*; Aug2010, Vol. 3 Issue 8, p. 521.

A.N Gusev, N.A. Zakharova, B.S. Frol

The computer technologies in the evaluation of voice characteristics and facial expressions

Abstract. The article considers the problem of improving the validity and reliability of the results of the psychological evaluation of behavior based on audio and video materials analysis. The principal benefits of joint use of modern computer systems such as The Observer XT, FaceReader and LVA-i for analysis of subject`s movements, expression of emotions on his/her face and voice. A brief overview of the functionality of these systems, allowing for qualitative and quantitative estimation of the real behavior of the test persons with the help of registration and subsequent analysis of a number of informative indicators of human behavior is presented. The results of the analysis of one case is presented: an interview with Yu.. The necessity of joint analysis of different behavioral indicators is emphasized.

Key words: psychological expertise of audio and video recordings, motor and verbal behavior, facial expressions, vocal parameters of the voice.