Выбор решений при распознавании эмоций по речи

В. П. Кальян
vkalyan@mail.ru
ФИЦ «Информатика и управление» РАН, Россия, г. Москва, ул. Вавилова, 44/2

Описывается выбор решений в системе распознавания эмоционального состояния человека по речи. Анализируется информативность измерительной базы распознавания на основании паралингвистических, артикуляционных и экстралингвистических особенностей речи с учетом индивидуальных эмоционально-смысловых коннотаций в речи испытуемого, описываются алгоритмы распознавания эмоций по речи, осуществляется выбор из множества решений и их верификация в отношении искренности и правдивости говорящего с учетом ситуативного контекста.

Ключевые слова: *распознавание эмоций; эмоциональная речь; древо принятия решений; пространство речевых признаков; паралингвистические особенности речи; артикуляционные модели; спектральная динамика; речевые форманты; частота основного тона; высота голоса*

DOI: 10.21469/22233792.2.4.07

1 Введение

Возможность определения эмоционального состояния говорящего по речи имеет большое практическое значение. Эмоции естественным образом сопутствуют речи, являясь особым каналом общения, по которому непосредственно передается отношение говорящего к текущей ситуации и содержанию сказанного. Это отношение невольно проявляется в характере речи — данное простое соображение, казалось бы, позволяет утверждать, что построение системы, связывающей параметры речи с ее искренностью и правдивостью, является непреложным фактом, а такие бесконтактные детекторы лжи скоро заменят существующие уже много лет в криминалистической практике полиграфы.

И действительно, уже более 40 лет на рынке услуг по детекции лжи под разными названиями появляются коммерческие приборы, позиционирующие сами себя как «анализаторы стресса в голосе». В 1972 г. в США был получен первый патент на прибор под названием PSE (Psychological Stress Evaluator), чуть позже другими разработчиками был выпущен в свет VSE (Voice Stress Evaluator). Утверждалось, что эти и подобные им приборы (Mark-II, ESM, Hagoth и др.) в отличие от полиграфа способны устанавливать неискренность без подключения к телу человека датчиков, а путем измерения изменений в голосе, обусловленных стрессом, который сопровождает ложные высказывания.

Суть работы анализаторов стресса в голосе объяснялась тем, что с помощью данных приборов якобы выделяются не воспринимаемые на слух акустические характеристики голоса, обусловленные стрессом. В частности, производители PSE утверждали, что этот прибор измеряет низкочастотную (около 10 Гц) модуляцию частоты основного тона голоса, обусловленную тремором мышц.

Проведенные экспертами-полиграфологами независимые исследования, исследования Американской ассоциации полиграфа (APA), а также тесты существующих на рынке приборов, проведенные Институтом полиграфа Минобороны США (DoDPI, показали, что точность этих приборов находится на уровне случайного угадывания [1].

Причину эксперты связывают с тем, что данные приборы лишь имитируют работу полиграфа. Они несравнимо беднее по возможностям измерения и анализа речевого сигнала классических систем распознавания речи, не учитывают опыт использования традиционного полиграфа с его обширной тактико-аналитической базой. В документации на эти приборы отсутствует описание принципов их работы под предлогом неразглашения коммерческой тайны, а инструкции применения отсутствуют или просто скопированы с инструкций некоторых моделей классического полиграфа.

Все это дискредитирует саму идею создания и использования речевого полиграфа на практике, требует постановки и решения задачи о применимости речевых анализаторов в оценке и интерпретации эмоциональных реакций человека по речи в части определения правдивости и искренности сказанного.

Для построения реально действующей системы распознавания эмоционального состояния говорящего по речи должны быть основательно проработаны измерительная база и алгоритмическая основа системы распознавания эмоций по речи.

Задача автоматического распознавания эмоциональной окрашенности звучащей речи является междисциплинарной и постоянно привлекает исследователей разных специальностей — не только лингвистов, но и математиков, программистов, психологов, физиологов [2].

Исследования ведутся по нескольким направлениям.

1. Модальность эмоций. Это традиционное направление работ психологов по изучению и классификации эмоций, выявлению эмоционально-смысловых коннотаций.
2. Нахождение объективных характеристик проявления эмоций в речи, связи эмоций с паралингвистическими, экстралингвистическими и артикуляционными особенностями речи. Одно из традиционных направлений работы лингвистов.
3. Способы извлечения эмоциональных характеристик из речевого сигнала. Построение пространства признаков для распознавания эмоций по речи. Работы ведутся смешанными коллективами, состоящими из физиологов, лингвистов, специалистов по автоматическому распознаванию речи.
4. Нахождение эффективных стратегий распознавания. Построение стратегий, алгоритмов и систем распознавания эмоций по речи. Верификация смыслов эмоциональных речевых реакций в зависимости от ситуативного контекста, выбор решений в отношении правдивости и искренности говорящего. Работы ведутся смешанными коллективами, состоящими из лингвистов, специалистов по автоматическому распознаванию речи, искусственному интеллекту.

По первому направлению можно выделить работы П. В. Симонова, К. В. Анохина, В. О. Леонтьева [3], К. Э. Изарда и A. R. Damasio, в которых были выделены группы эмоций. Среди этих групп принято выделять первичные и вторичные.

Первичные эмоции считаются базовыми, врожденными. Они включают в себя обобщенные, близкие к рефлексу («автоматические», или запрограммированные), страх и мгновенные реакции на стимулы, представляющие опасность. Они не предполагают сознательных размышлений и включают в себя шесть базовых эмоций, выделенных Дарвином: страх, гнев, отвращение, удивление, грусть и счастье; впрочем, по К. Изарду выделяют 11 фундаментальных (базовых) эмоций: радость, удивление, печаль, гнев, отвращение, презрение, горе-страдание, стыд, интерес-волнение, вина, смущение.

Вторичные эмоции — это более сложные эмоции, и они задействуют высшие центры коры головного мозга. Они могут заключать в себе базовые эмоции гнева или страха или

Рис. 1 Пример непрерывной шкалы эмоций

иметь более сложную структуру, например к ним добавятся сожаление, тоска, стыд, вина, зависть или ревность. Вторичные эмоции не являются автоматическими: они производятся мозгом, индивид думает о них и принимает решение, что с ними делать — какие действия лучше всего предпринять в той или иной ситуации.

Сознательные размышления и вторичные эмоции влияют на то, как индивид реагирует на ситуации, которые порождают первичные эмоции: он может отступить или смутиться предположив некоторую опасность, но, придя в себя, распознав и иначе оценив ситуацию, может, например, сделать вид, что ничего не случилось.

Главная проблема в обнаружении эмоционального состояния человека состоит в том, что все люди по-разному выражают свои эмоции. Кроме того, очень важно учитывать тонкие речевые компоненты и их изменение в процессе разговора. Поэтому исследователи от дискретной классификации эмоции и отнесения исследуемого фрагмента к какой-либо строго определенной категории эмоционального состояния переходят к описанию непрерывного эмоционального пространства, например к такому, как показано на рис. 1.

Преимущество такого подхода заключается в возможности выражать огромное количество эмоций: от «средней раздраженности» до «ярого гнева», — а также различать неуловимые отличия между очень схожими эмоциями.

Четырехмерную сферическую модель эмоций предложила группа исследователей в публикации В. А. Вартанова [4]. Построение модели проводилось экспериментально с помощью многомерного шкалирования субъективных различий между эмоциональными состояниями, задаваемыми специально созданными образцами. Чтобы уровнять и сделать определенным содержание этих образцов, в эксперименте использовалось одно и то же слово, произнесенное в разных эмоциональных состояниях. В одной серии использовалось слово «да», а в другой — «нет». Полученные параметры (факторы) характеризовались как бимодальные спектральные фильтры. Из них выделили четыре измерения: лучше–хуже, удивление–уверенность, симпатия–равнодушие, активное–пассивное отвержение.

По второму направлению лингвисты и психологи выявляют эмоциональные составляющие речи, анализируя ее паралингвистические, экстралингвистические и артикуляционные особенности.

Из прикладной лингвистики и апеллятивной фонетики известно, что многие признаки эмоционального состояния, искренности и правдивости говорящего содержатся в мело-

дике, акцентуации, смене темпа и ритма речи, особенностях артикулирования, дрожании голоса — особенно в стрессовых ситуациях, например при ответах собеседника на неожиданные «неудобные» вопросы.

Например, некоторыми исследователями установлена связь направления движения высоты голоса с положительными или отрицательными эмоциями: понижение высоты — с приятными эмоциями, а ее повышение соотносят с удивлением или страхом. Большое значение придают специалисты завершающему фрагменту мелодического контура фразы, поскольку он может информировать не только о повествовательном, вопросительном или восклицательном типе предложения, но и об отношении говорящего к теме высказывания, ситуации общения, к собеседнику.

Эмоциональную составляющую речитации обнаруживают не только в просодии (мелодике, ритме, акцентуации, темповой и ритмической динамике речи), но и в артикулировании (характере произнесения гласных, согласных, слогов, слов). Данные апеллятивной фонетики содержат важную информацию об эмоциональном состояния говорящего, что позволяет использовать артикуляционные модели в задачах распознавания эмоций по речи [5]. Экстралингвистические особенности речи проявляются в дрожании голоса, паузах, придыхании, заикании, покашливании, смехе [6]. Перед исследователями стоит задача установить их эмоциональную обусловленность, соотнеся с описанной выше модальностью эмоций.

По третьему направлению — основная задача получения признаков эмоциональной составляющей речи состоит в том, чтобы преобразовать звуковую волну в такое признаковое пространство, в котором множество объектов одного класса будет сгруппировано вместе, а множество объектов альтернативных классов максимально разнесено. Из всего спектра работ на современном этапе можно выделить четыре группы объективных признаков и соответствующих методов, позволяющих различать речевые образцы: спектрально-временны́е, кепстральные, амплитудно-частотные и признаки на основе нелинейной динамики [4–9].

Четвертое направление — нахождение эффективных механизмов и стратегий распознавания для создания речевого полиграфа; построение алгоритмов, сценариев и, наконец, систем распознавания правдивости и искренности говорящего по речи [10]; верификация смыслов эмоциональных речевых реакций в зависимости от ситуативного контекста; выбор решений.

Нельзя не признать, что одни и те же феномены эмоциональной речи в зависимости от ситуации могут быть интерпретированы по-разному. При маркировке тех моментов в высказывании, где проявляется волнение, в контексте ситуации может быть учтено влияние обстоятельств на общую картину эмоций и смысл происходящего в момент речевого высказывания [11].

Например, в зависимости от ситуации нескрываемый гнев говорящего (проявляется, например, в характерной смене темпа и ритма речи, тщательном выговаривании согласных в словах) может свидетельствовать о неверно высказанном предположении, содержащемуся в вопросе, заданном испытуемому, или о его отношении к самой ситуации допроса, к допрошивающему; растерянность и смущение, проявляющиеся в неуверенной речи, могут говорить как о страхе разоблачения, так и о непонимании вопрооса. Дрожание голоса в зависимости от ситуации может свидетельствовать об обиде, страхе, гневе или, наоборот, радости.

Для сопоставления речевых реакций с ситуацией, в которых они проявлялись, нами был разработан нормативный язык описания модели ситуаций, опирающийся на наше понимание их морфологии [11].

Настоящая работа посвящена экспериментам по созданию системы распознавания правдивости и искренности говорящего.

2 Постановка задачи

Перед началом эксперимента по автоматическому или экспертному распознаванию правдивости и искренности говорящего по речи исследователи в качестве исходных данных, как правило, располагают:

- речевым сигналом, представленным в виде дискретной функции от времени — последовательности временны́х отсчетов $C(k)$, где k — номера отсчетов сигнала по оси времени с фиксированным шагом;
- информацией о характере говорящего, т. е. о характерных для него когнитивных, регулятивных и коммуникативных особенностях проявления эмоций E;
- первичной характеристикой расследуемой ситуации в виде совокупности обстоятельств, описанных некоторым нормативным языком G.

Перед нами стоит задача выявления эмоциональной составляющей речи по паралингвистическим, экстралингвистическим, артикуляционным особенностям высказывания и распознавание смысла эмоции индивида в контексте ситуации дознания. Эти эмоции, являясь непроизвольной реакцией индивида на попытку исследователя тем или иным образом прояснить исследуемую ситуацию, должны были бы позволить сделать заключения в отношении сказанного испытуемым и прояснить как позицию испытуемого в ситуации дознания, так и общую картину происшествия.

Однако неоднозначность эмоционально-смысловых коннотаций в проекции на реконструируемую картину происшествия может привести к существенным ошибкам и делает необходимой выработку специальной стратегии для выбора решений в распознавании смысла речевых эмоций.

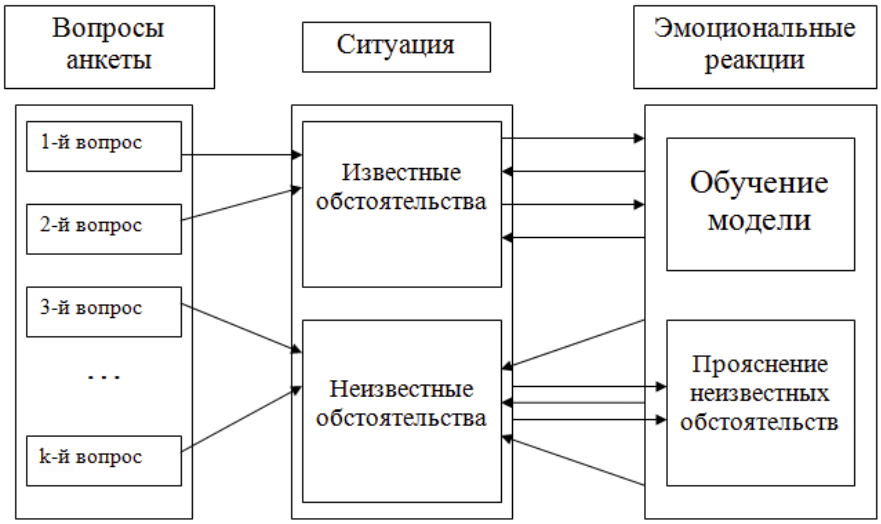

Рис. 2 Обучение модели и распознавание эмоциональной реакции человека в ситуативном контексте

В данной работе мы исходим из предположения о том, что верификация смыслов эмоциональных проявлений в речи по выработанным признакам становится возможной с помощью сопоставления эмоционально-смысловых коннотаций с ситуативным контекстом при условии применения специальных процедур опроса испытуемого в процессе реконструкции расследуемого события (рис. 2).

3 Описание эксперимента

В настоящей работе распознавание эмоций и заключение о правдивости и искренности испытуемого опиралось на:

- паралингвистические особенности речи (т. е. ее мелодику, акцентуацию, темпоритм, см. рис. 3–5), характерные для индивида;
- индивидуальные особенности артикулирования;
- экстралингвистические особенности высказывания; к ним относятся — паузы, смех, покашливание, вздохи, плач, мычание, заикание, дрожание голоса;
- знания об эмоционально-смысловых коннотациях, характерных для речи испытуемого;
- соотнесение эмоциональности высказывания с ситуативным контекстом.

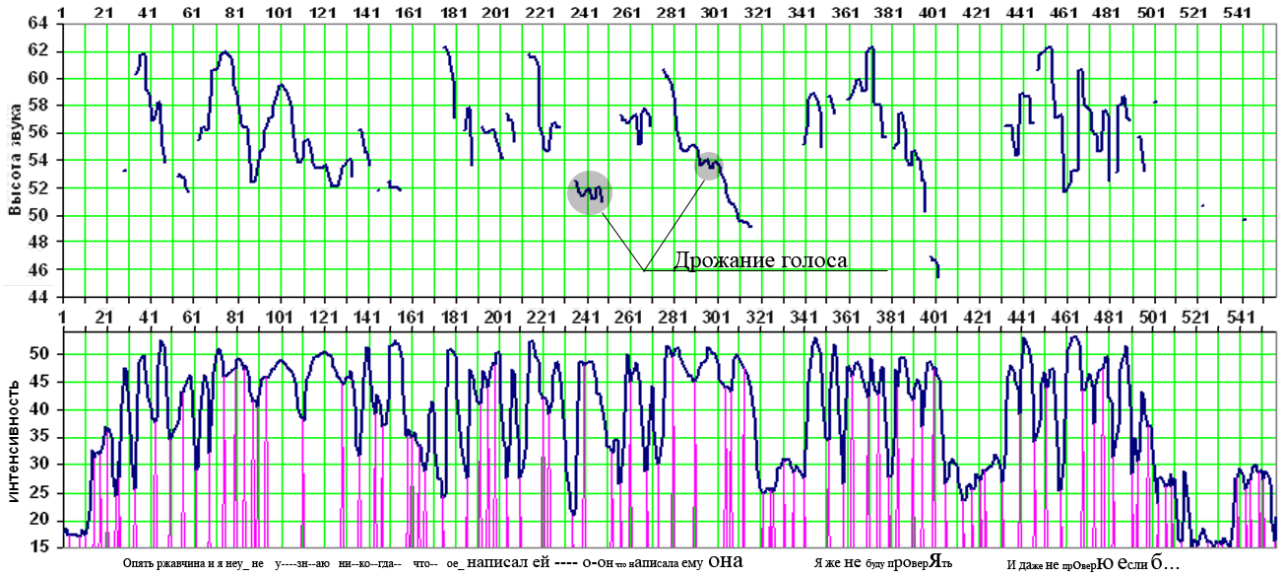

Рис. 3 Первичная сегментация по минимумам интенсивности звука в высказывании «что написал...» и характерные фрагменты дрожания голоса в словах «он», «она»

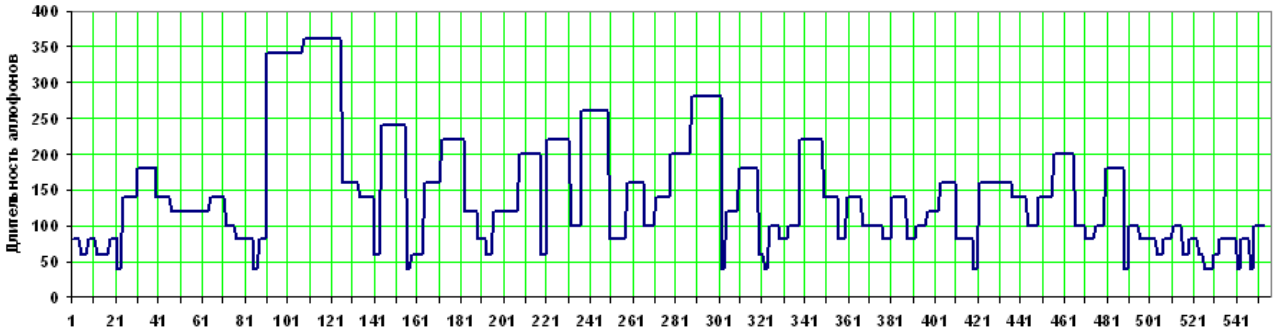

Рис. 4 Положение длительностей аллофонов в высказывании «Что написал»

Рис. 5 Динамика темпа произнесения гласных

3.1 Измерительная база

На первом этапе обработки данных речевой сигнал $C(k)$ подвергался спектральному анализу посредством быстрого преобразования Фурье (БПФ) с последовательно сдвигаемым взвешенным окном. Вычислялся динамический спектр в виде последовательности значений кратковременных энергетических спектров $S(w, i)$, измеренных в моменты времени каждые 20 мс, траектории максимумов трех первых формант $F(j, i)$, кривые интенсивности в низком, среднем и высоком частотных диапазонах $F(l, i)$ — так называемая «гребенка», амплитудная огибающая общей интенсивности $A(i)$ и звуковысотный контур речевой просодии $P(i)$, рассчитав для этого по специальным алгоритмам траекторию основного тона.

На основании этих данных:

- произвели первичную сегментацию по минимумам интенсивности звука (см. рис. 3);
- произвели маркировку аллофонов $A(m)$; сегменты идентифицировали по их спектральных характеристикам и на основании справочных материалов или экспертных оценок по типу аллофон гласного/согласного (вокализованный, щелевой, взрывной и т. п.);
- скорректировали, перегруппировали первичную сегментацию и вычислили длительности звуков, соответствующих гласным и согласным (рис. 6);

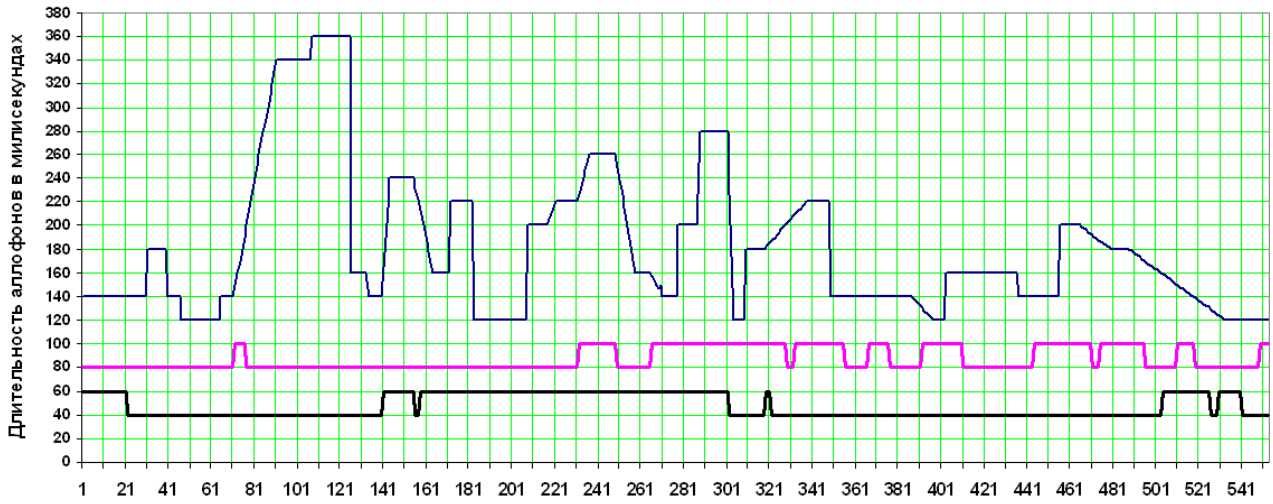

Рис. 6 Графики динамики длительностей гласных (верхний), сонорных, щелевых (средний) и взрывных (нижний) согласных в высказывании «что написал...»

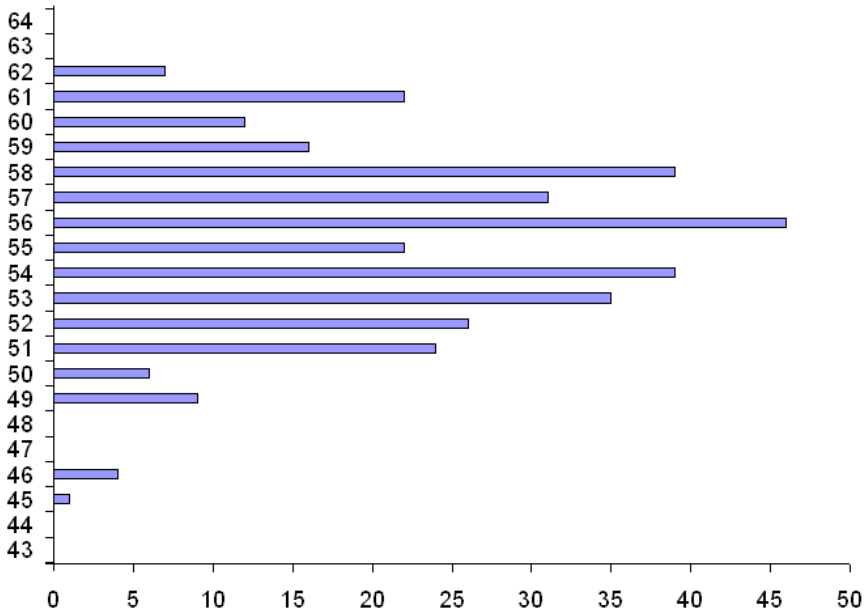

Рис. 7 Гистограмма высоты голоса в высказывании «что написал...»

– выявили просодию, т. е. паралингвистические особенности речи, как то:
  - высоту голоса привели к непрерывной музыкальной шкале стандарта MIDI, где нота «Do» первой октавы соответствует 52-м, «Re» — 54-м и т. д., по этой шкале анализировали мелодику речевого высказывания;
  - динамический темп речи (см. рис. 5);
– распознали ритмические формы;
– определили особенности интонирования, например установили присутствие элементов контрастно-регистрового интонирования, что наглядно видно на рис. 3 и 7 во временно́м диапазоне 273–321 отсчетов: присутствует бросок высоты голоса на октаву вниз менее чем за 1,5 с.

Полученные данные использовались:

– на этапе обучения модели при экспертной разметке на эпизоды, свидетельствующие об эмоциональности речи и выявлении характерных для индивида эмоционально-смысловых коннотаций;
– на этапе распознавания для выявления эмоциональной составляющей речи, заключения об искренности говорящего и правдивости сказанного.

3.2 Выявление признаков эмоций в речи

В результате серии опытов и экспертных заключений об эмоциональности речевых фрагментов наиболее информативными оказались следующие признаки:

– длительность ударных и безударных гласных по отношению к средней длительности их произнесения в текущем эпизоде позволяют выявить фразовые и эмоциональные акценты;
– удлинение предударных щелевых или сонорных согласных является средством эмоционального усиления акцента говорящим;
– преувеличенная акцентуация ударного слога в слове за счет интенсивности звука голоса — свидетельство эмоционального возбуждения;

– акцентуация за счет увеличения длительности ударных гласных и предударных согласных — свидетельство желания убедить собеседника;
– обратная величина динамики длительности гласных дает динамику темпа речи (см. рис. 5);
– изменение темпа речи на уровне слова, фразы, высказывания свидетельствует об осмысленном проявлении отношения говорящего к смыслу высказывания, желании выделить или скрыть это отношение;
– эмоциональное усиление акцента в слове, фразе часто сопровождается «двойной акцентуацией» внутри ударных гласных, появляются два максимума на кривой интенсивности, которые состоят из двух сегментов первичной разбивки;
– речевой ритм, построенный на соотношении длительностей соседних ударных и безударных гласных, помимо того что позволяет распознавать акцентуацию, уточнять наличие структурирующих (словесных, фразовых) и эмоциональных акцентов, выявляет мультипликативные формы (например, скандирование), чаще всего имеющие эмоциональную природу;

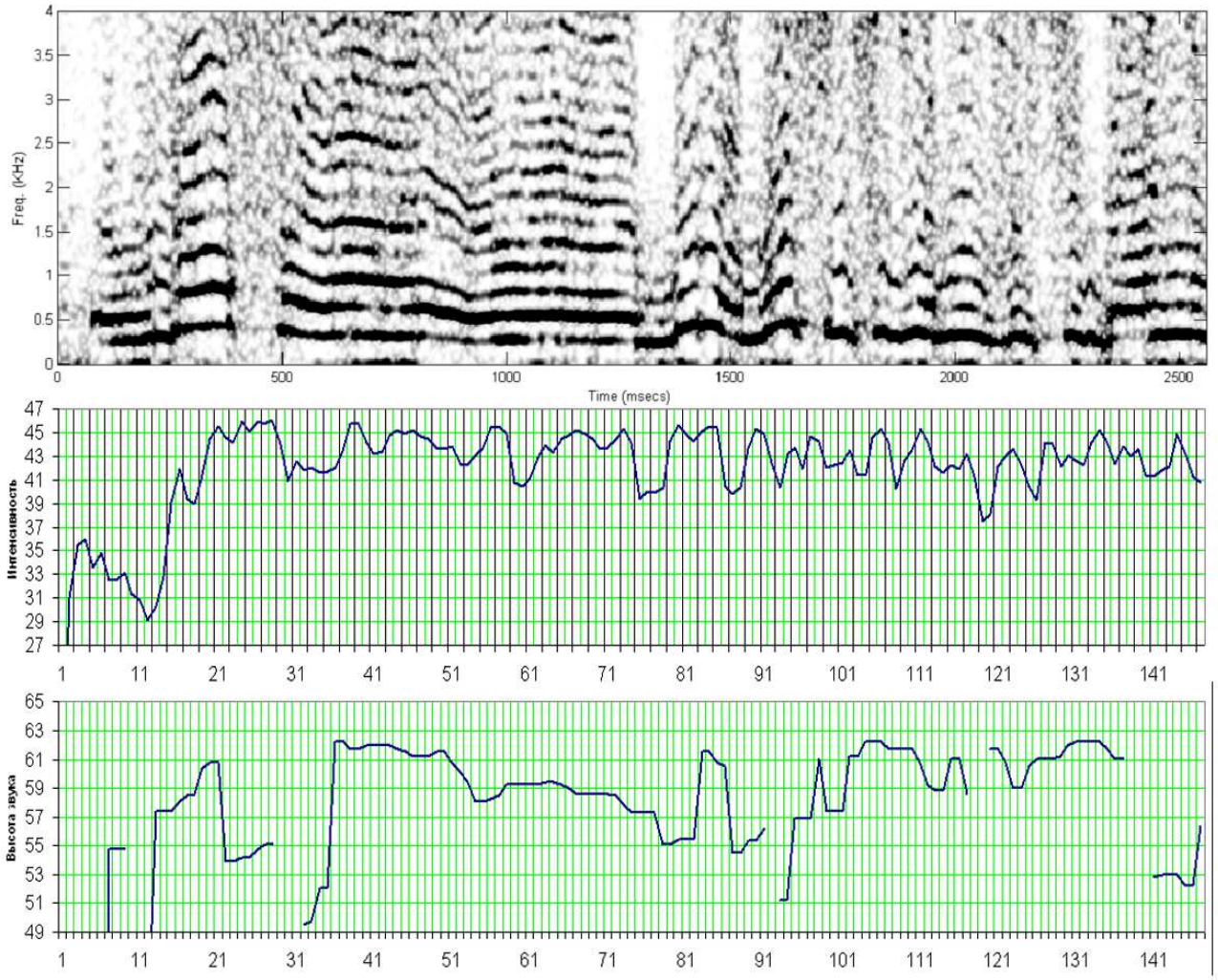

Рис. 8 Интенсивность и высота звука в высказывании «У нас э-э-э-э было восемь судебных заседаний»

– дрожание голоса (относится к экстралингвистическим элементам речи, см. рис. 3) свидетельствует о непроизвольно проявляющемся волнении — чаще всего от негодования, страха, обиды или, наоборот, от радости, восторга; и если негодование, радость и восторг обычно сопровождаются повышенным уровнем интенсивности звука, то страх и обида проявляются средним или пониженным уровнем интенсивности;
– характерное периодическое чередование взрывных участков и пауз (смычек) свидетельствует о смехе, покашливании в речи;
– длительные (порядка секунды и более) вокализованные «А», «Э», «М» свидетельствуют о неуверенной речи, неподготовленности речевого высказывания.

На рис. 8 изображен фрагмент амплитудной и высотной характеристик типичного неуверенного «блеяния» другого диктора во фразе «У нас э-э-э-э было восемь судебных заседаний».

На основании предварительных данных о проявлении эмоций в речи в виде набора эмоциональных речевых признаков $\boldsymbol{me}$ мы построили, а впоследствии верифицировали с помощью экспертных оценок характерный для испытуемого индивида набор эмоционально-смысловых коннотаций.

В результате наших исследований [5, 6, 8] были выявлены ранее неизученные связи речевых признаков с модальностью эмоций, такие как:

– контрастно-регистровое интонирование, означающее испуг, панику (см., например, октавный бросок в траектории высоты голоса в слове «она» на рис. 3);
– смена ритма со сложного на простой, означающая раздражение, гнев;
– двойная акцентуация гласных, означающая возмущение;
– подмена гласных в акцентируемом слоге, например «а» на «ы» (пример во фразе «сама понимаешь» первая гласная «а» звучит как «ы»), свидетельствует об агрессивности, гневе, злости, возмущении; вычисляется по артикуляционным моделям гласных на основании значений первых трех формант в распознаваемом сегменте (рис. 9–11).

Кроме того, были приняты во внимание взаимозависимости высоты голоса, интенсивности звука, темпа, разборчивости и уверенности речи, полученные другими исследователями [4, 7, 9, 12]:

– явно высокий звук — энтузиазм, радость, испытуемый заинтересован и проявляет интерес;

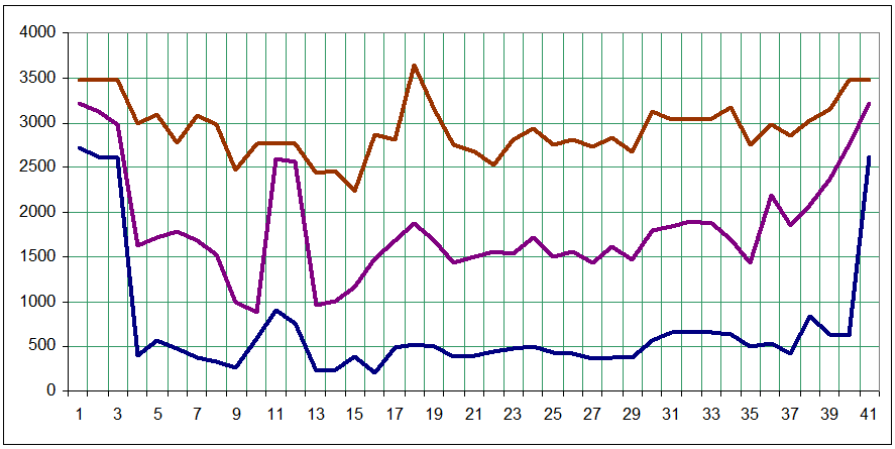

Рис. 9 Траектория трех первых формант во фразе «сама понимаешь»

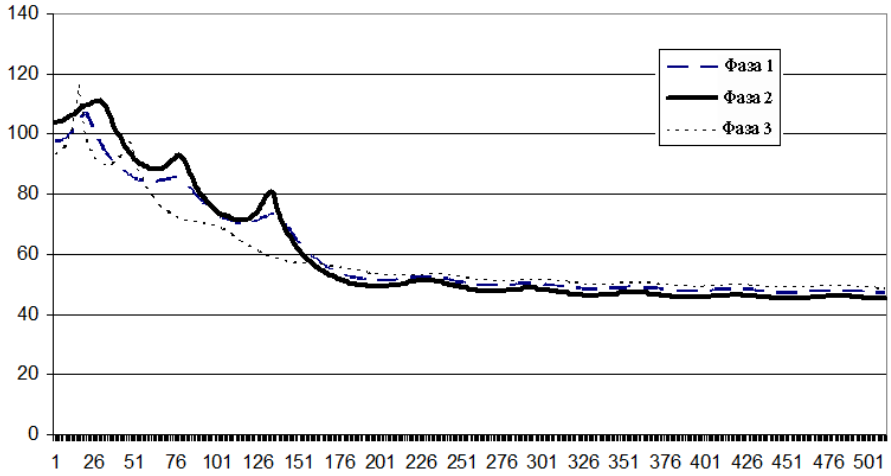

Рис. 10 Фазы LPC-огибающей динамики спектра произнесения гласной в первом слоге слова «сама»

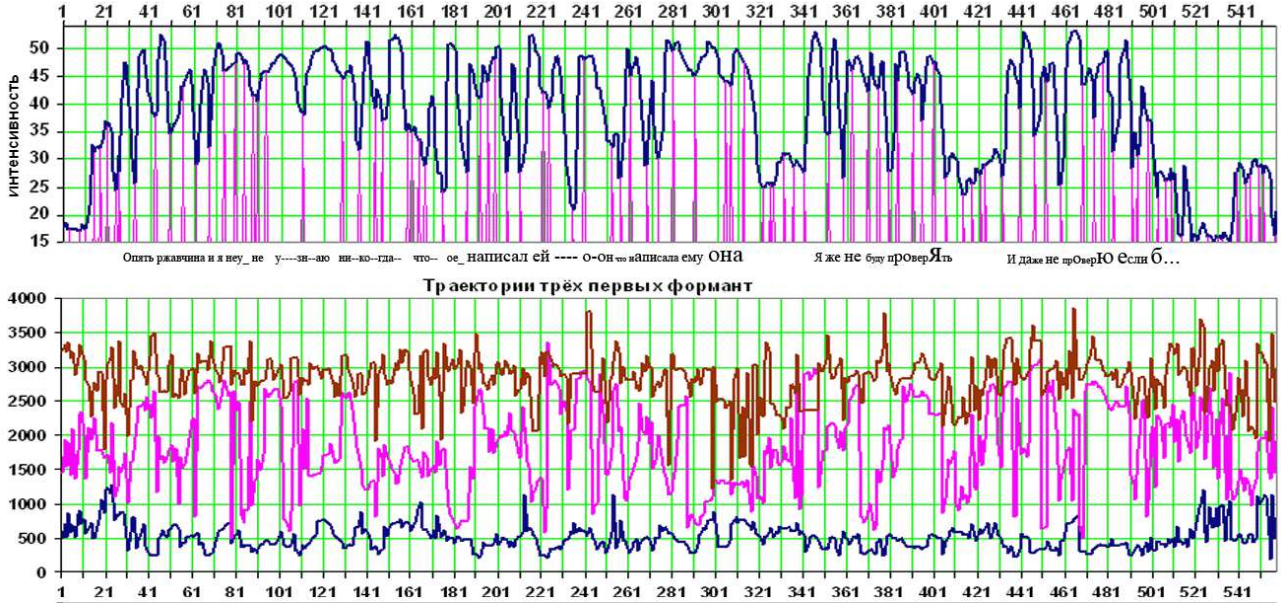

Рис. 11 Графики интенсивности и траектории трех первых формант для выявления характера произнесения гласных

- чрезмерно высокий, пронзительный — беспокойство;
- мягкий и приглушенный, с понижением интонации к концу каждой фразы — печаль, усталость;
- форсирование звука — напряжение, обман;
- быстрая речь — очевидная взволнованность — желание убедить или уговорить кого-то;
- медленная речь — высокомерие, усталость, угнетенное состояние;
- прерывистая речь – неуверенность;
- лаконичность и решительность речи — явная уверенность;
- заикание — напряженность или обман;
- нерешительность в подборе слов — неуверенность в себе или намерение внезапно удивить чем-то;

– появление речевых недостатков (повторение или искажение слов, обрывание фраз на полуслове) — несомненное волнение, но иной раз и желание обмануть;
– опускание речевых пауз — напряжение;
– слишком удлиненные паузы — незаинтересованность или несогласие.

Для создания модели реакций испытуемого был выделен специальный этап исследования — создание массива данных, где накапливались данные о речевых реакциях испытуемого, проводилось выделение значимых параметров.

3.3 Построение и обучение модели

Для разбиения текущих значений речевых параметров на классы (группы признаков) таких непрерывных параметров, как мелодический контур, огибающая интенсивности речевого сигнала или динамика темпа речи, длительность пауз, использовался метод кластеризации, вероятностный подход. Предполагалось, что каждый рассматриваемый на этапе обучения объект (эмоциональная речевая реакция) относится к одному из k классов обучающих выборок. Для определения центроида кластера вычислялась медиана и производилось обучении модели.

Принадлежность сегмента определялась после вычисления его метрики по группе указанных выше параметров в соотнесении с алфавитом сегментов, выявленных и маркированных в процессе экспертной оценки на этапе обучения модели; его многомерная классификация и соответственно маркировка осуществлялись при выполнении ряда условий.

Обозначим множество темпорально-акустических характеристик речевого высказывания как $\boldsymbol{M}$ (из которых подмножество $\mathbf{me}$ свидетельствует об эмоциональной окраске, так что $\mathbf{me}(\boldsymbol{i})$ — набор признаков эмоций в речи, где $i = 1, \ldots, N$, и N — количество классов эмоциональной окраски, отражающихся в речевых параметрах), совокупность эмоционально-смысловых коннотаций индивида как $\boldsymbol{E}$, морфологию ситуаций дознания и реконструируемого происшествия как упорядоченные множества $\boldsymbol{G_1}$ и $\boldsymbol{G_2}$.

В данной работе была поставлена задача выбора решений об искренности говорящего и правдивости сказанного им при распознавании эмоций по речи в связанной системе $\boldsymbol{M, E, G_1, G_2}$. Перед проведением эксперимента была построена модель возможных эмоциональных реакций испытуемого $\boldsymbol{G_1}$–$\boldsymbol{G_2}$–$\boldsymbol{M}$–$\boldsymbol{E}$, которая во время эксперимента обучалась по примерной схеме, отображенной на рис. 2.

Итак, у нас есть упорядоченное подмножество $\mathbf{me}$ множества $\boldsymbol{M}$ признаков, распознанных экспертами как характеристики, которые свидетельствуют о волнении говорящего или желании диктора выделить слово, артикулируя звуки в нем особым образом. У нас это множество разбито на N классов и представлено массивом данных $\mathbf{me}(k, i, m, L)$, где каждый k-й элемент из $\boldsymbol{M}$ отнесен к i-му классу и каждому классу поставлено в соответствие значения из $\boldsymbol{S}$ — группы частично упорядоченных параметров, представленных массивом $\boldsymbol{S}(i, m, c, d)$, где i — имя (номер) класса; m — имя (номер) параметра; c — положение центроида класса; d — медиана класса i.

Тогда $\boldsymbol{\Delta\widehat{L}}$ — ближайшее расстояние между кластерами значений (векторная разность) параметров i-х классов массива $\boldsymbol{S}$ и соответствующими значениями параметров распознаваемого $(n+1)$-го сегмента в пространстве признаков, т. е.

$$\Delta\widehat{L} = \arg\min_i [\widehat{M}(k+1) - \widehat{S}(i)].$$

Эмоциональная окраска сегмента может рассматриваться как вероятность, вытекающая из величины отклонения его параметров от некоторых «нормальных» значений

для данного контекста. Здесь мы опираемся на имеющиеся в наличии аналогичные по контексту артикуляционные позиции, которые могут быть распознаны как сегменты, соответствующие «спокойному артикулированию».

Для фиксации наличия признаков эмоций в речи, таких как смена ритма со сложного на простой (которая устанавливалась с помощью автокорреляционной функции длительностей гласных), контрастно-регистровое интонирование (которое определялось на плоскости по расстоянию между пиками функции плотности вероятности высоты голоса и временно́й дистанции между их значениями в речевом фрагменте) и т. п. выявлялся сам факт наличия такого признака, т. е. использовалась бинарная оппозиция — есть, нет (true, false).

3.4 Пример выбора решения

В выборе решения о подлинном смысле и правдивости сказанного учитывалось соотнесение трёх групп признаков:

- акустико-временны́х — тональных, спектродинамических и темпоральных характеристик речи и вычисленных по ним данных о просодии и артикуляции высказывания;
- эмоционально-смысловых коннотаций речи;
- ситуативных — данных о ситуативном контексте высказываний; при этом рассматривалась морфология двух разных, но связанных между собой ситуаций — текущая ситуация дознания и модель реконструируемой следствием цепи событий.

Здесь вопросы дознания к испытуемому на этапе обучения модели наряду с дискретными состояниями описания ситуации дознания, которые относятся ко множеству обстоятельств $\boldsymbol{G_1}$, задаются по известным обстоятельствам как $\boldsymbol{G_1}$, так и $\boldsymbol{G_2}$. По реакции индивида на вопросы по заранее известным обстоятельствам происходит обучение модели распознавания.

При накоплении достаточной представительности обучаемой модели испытуемому задаются вопросы в отношении неизвестных следствию обстоятельств $\boldsymbol{G_2}$, эмоциональные речевые реакции $\mathbf{me}$ испытуемого на вопросы из множества $\boldsymbol{G_2}$ соотносятся со множеством смысловых коннотаций $\boldsymbol{E}$, на основании чего делается вывод об искренности и правдивости ответа. При этом информативными оказываются как искренние, правдивые ответы, так и ложные, так как они свидетельствуют о попытке сокрытия обстоятельств, которые нужны для дополнения описания $\boldsymbol{G_2}$. В этом случае в части модели ситуации дознания $\boldsymbol{G_1}$ может быть сформирован дополнительный сценарий для прояснения обстоятельств, которые пытался скрыть испытуемый.

Анализируя исходное речевое высказывание $C(k)$, касающееся описания ситуции $\boldsymbol{G_2}$ в ряду темпорально-акустических признаков $\boldsymbol{M}$, мы выделяем из перечисленных 28 признаков 8 значимых для данного высказывания и устанавливаем эмоционально-смысловые коннотации $\boldsymbol{E}$, связанные с элементами исследуемой ситуации $\boldsymbol{G_2}$, одновременно выявляя неоднозначности эмоционально-смысловых коннотаций.

Так, например, увеличение темпа речи во временно́м диапазоне 125–210 отсчетов на рис. 3 и снижение темпа в диапазоне 210-250 отсчетов может свидетельствовать как о желании убедить собеседника, так и о неуверенности говорящего, его волнении.

Природа волнения становится понятной из сопоставления с другими эмоциональными признаками в этом же речевом фрагменте — из дрожания голоса на указанном участке снижения темпа на слове «он» и следующим за ним фрагменте (273–321 отсчетов) с эмоциональными признаками — контрастно-регистровом интонирования и дрожании голоса на

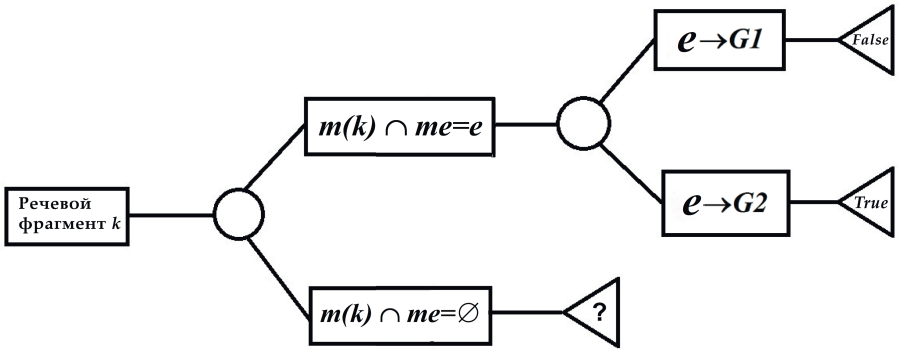

Рис. 12 Схема вычисления целевой переменной на основании акустико-темпоральных и ситуативных признаков

слове «она» с двойной акцентуацией на последнем слоге этого слова, свидетельствующими о страхе и глубокой обиде.

Ситуативный контекст проясняет смысл эмоционального всплеска — здесь речь идет о подозрении в супружеской неверности. Функция испытуемого — пострадавший.

Соотнесение многофакторных признаков и заключение об искренности и правдивости говорящего осуществлялось с помощью дерева принятия решений. Структура дерева представляла собой «листья» и «ветки». На ребрах («ветках») дерева решения записывались атрибуты (речевые признаки $\boldsymbol{M}$ и ситуативный контекст $\boldsymbol{G_1, G_2}$), от которых зависела целевая функция (смысл эмоциональной реакции), в «листьях» были записаны значения целевой функции, а в остальных узлах — атрибуты, по которым разветвляется дерево и из которых одна из ветвей приводит к конечному значению целевой функции.

Таким образом, для классификации очередного фрагмента речевой реакции нужно спуститься по дереву до листа и выдать соответствующее значение.

Эта схема отражает модель, которая вычисляет значение целевой переменной на основе нескольких акустико-темпоральных и ситуативных переменных параметров (признаков) на входе (рис. 12).

По указанной схеме в процессе эксперимента был реализован алгоритм распознавания в связанной системе $\boldsymbol{M, E, G_1, G_2}$ — от последовательности эмоционально значимых речевых фрагментов и алфавита эмоционально-смысловых коннотаций до процессов дознания и реконструкции исследуемой ситуации.

4 Заключение

В работе описан опыт выбора решений в системе распознавания эмоционального состояния человека по речи. Анализ эмоциональных проявлений на основе соотнесения пара- и экстралингвистических особенностей и артикуляционных моделей речи с их эмоционально-смысловыми коннотациями показывает неоднозначность этих коннотаций, что в проекции на реконструируемую картину происшествия и ситуацию дознания может привести к существенным ошибкам распознавания. Предложена стратегия выбора решений распознавания эмоционального состояния человека по речи в связанной системе темпорально-акустических, эмоционально-смысловых и ситуационных зависимостей. При настоящем подходе верификация смыслов эмоциональных речевых реакций становится возможной благодаря именно сопоставлению с ситуативным контекстом.

Литература

[1] *Князев В., Варламов Г.* Полиграф и его практическое применение. — Принт-Центр, 2012. 859 с.

[2] *Кальян В. П.* Музыка, речь и компьютер. — М.: ВЦ РАН, 1998. 38 с.

[3] *Леонтьев В. О.* Десять нерешенных проблем теории сознания и эмоций. — Одесса, 2008. http://polatulet.narod.ru/dvc/com/vleontiev_problems.html.

[4] *Вартанов А. В.* Антропоморфный метод распознавания эмоций в звучащей речи // Национальный психологический ж., 2013. № 2[10]. С. 69–79. http://www.psy.msu.ru/science/npj/journals/npj-no10-2013.pdf.

[5] *Кальян В. П.* Исследование применимости артикуляционных моделей в задачах распознавания эмоций по речи // Докл. 9-й Междунар. конф. «Интеллектуализация обработки информации». — М.: ТОРУС ПРЕСС, 2011. С. 334–349.

[6] *Кальян В. П.* Построение алгоритмов распознавания эмоционального состояния человека по пара и экстралингвистическим особенностям речи // Модели и методы распознавания речи. — М.: ВЦ РАН им. А. А. Дородницына, 2010. С. 24–46.

[7] *Schuller B., Steidl S., Batliner A.* The INTERSPEECH 2009 emotion challenge // Interspeech, 2009. Т. 2009. С. 312–315. http://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_0312.pdf.

[8] *Кальян В. П.* Разработка алгоритмов распознавания эмоционального состояния человека по паралингвистическим особенностям речи // Докл. 15-й Всеросс. конф. «Математические методы распознавания образов». — М.: МАКС-Пресс, 2011. С. 334–349.

[9] *Брестер К. Ю.* Коллективный эволюционный метод многокритериальной оптимизации в задачах анализа речевых сигналов. Дисс. ... канд. техн. наук. — Красноярск: 2013. 143 с. http://research.sfu-kras.ru/sites/research.sfu-kras.ru/files/Dissertaciya_Brester_K.Yu_.pdf.

[10] *Кальян В. П.* Архитектура системы распознавания эмоционального состояния человека по речи // Модели и методы распознавания речи. — М.: ВЦ РАН им. А. А. Дородницына, 2013. С. 89–98.

[11] *Кальян В. П.* Морфология ситуации в системе распознавания эмоционального состояния человека по речи // Модели и методы распознавания речи. — М.: ВЦ РАН им. А. А. Дородницына, 2012. С. 92–102.

[12] *Сидоров К. В., Филатова Н. Н.* Анализ признаков эмоционально окрашенной речи // Известия ЮФУ. Технические науки, 2012. Т. 134. № 9. С. 39–45. http://eprints.tstu.tver.ru/69/1/3.pdf.

Поступила в редакцию 13.09.2016

Decision support in process of recognizing emotion in speech

V. P. Kalyan

vkalyan@mail.ru

Federal Research Center "Computer Science and Control" of RAS

44/2 Vavilova Str., Moscow, Russia

Background: Commercial devices, presenting themselves as "analyzers of stress in a voice," have been appearing in the market for lie detection services for more than 40 years. Those devices, unlike polygraphs, were claimed to be capable of establishing insincerity without requiring any connection to human body via sensors, but by measuring changes in one's voice

caused by raised stress level that is provided by making false statements. The independent researches of the devices existing in the market, conducted by polygraphology experts, American Association of a Polygraph (MACAW), Institute of Polygraph Tests of the USA Ministry of Defence (DoDPI), proved that the accuracy of those devices drops to the level of random guessing.

Methods: This work describes the experience of decision making in the system designed to recognize the emotional state of a person by his/her speech, concerning truthfulness and sincerity of what is being said. The information value of the recognition-measuring base is analyzed on the basis of paralinguistic, articulation, and extralinguistic speech features, regarding also individual emotional and semantic connotations of the testee's speech and the algorithms helping to recognize emotions by speech are described. We make the choice from a number of decisions and verify them regarding to the speaker‘s sincerity and truthfulness and concerning situational context as well.

Results: The analysis of emotional expressions based on matching para- and extralinguistic features and articulatory model of speech to their emotional and semantic connotations shows certain ambiguity of those connotations. It can lead to serious essential mistakes while recognizing if projected to the reconstructed accident picture and a situation of inquiry. A strategy for choosing decisions for identifying one's emotional state by his speech is proposed within within a related system of temporal and acoustic, emotional and semantic and situational dependences. This way gives one an opportunity to verify the meanings of emotional speech reactions due to correlating with the situational context.

Keywords: *recognition of emotions; emotional speech; decision-making tree; space of speech signs; paralinguistic features of speech; articulation models; spectral dynamics; speech formant; sound pitch; sound altitude*

DOI: 10.21469/22233792.2.4.07

References

[1] Kniazev, V., and G. Warlamov. 2012. *Poligraph i ego prakticheskoe primenenie* [Polygraph and its practical application]. Print-Center. 859 p.

[2] Kalyan, V. P. 1998. *Musyka, rech' i komp'yuter* [Music, speech, and computer]. Moscow: A. A. Dorodnitsyn CC RAN. 38 p.

[3] Leontiev, V. O. 2008. *Desyat' nereshennykh problem teorii soznaniya i emotsiy* [Ten unsolved problems in the theory of consciousness and emotions]. Odessa. Available at: `http://polatulet.narod.ru/dvc/com/vleontiev_problems.html` (accessed April 7, 2017).

[4] Vartanov, A. V. 2013. Antropomorfnyy metod raspoznavaniya emotsiy v zvuchashchey rechi [Anthropomorphic method of emotion recognition in sounding speech]. *Natsyonalnyi psikhologitcheskiy zh.* [National Psychological J.] 2(10):69–79. Available at: `http://www.psy.msu.ru/science/npj/journals/npj-no10-2013.pdf` (accessed April 7, 2017).

[5] Kalyan, V. P. 2012. Issledovanie primenimosti artikulyatsionnykh modeley v zadachakh raspoznavaniya emotsiy po rechi [Study of the applicability of articulatory models in speech recognition problems by speech]. *9th Conference (International) on Intellectualization of Information Processing Proceedings*. Moscow: TORUS PRESS. 498–502.

[6] Kalyan V. P. 2010. Postroenie algoritmov raspoznavaniya emotsional'nogo sostoyaniya cheloveka po para i ekstralingvisticheskim osobennostyam rechi [The construction of algorithms for recognizing the emotional state of a person according to para- and extralinguistic features of speech]. *Modeli i metody raspoznavaniya rechi* [Models and methods of speech recognition]. Moscow: A. A. Dorodnitsyn CC RAS. 24–46.

[7] Shuller, B., S. Steidl, and A. Batliner. 2009. The INTERSPEECH 2009 emotion challenge. *Interspeech* 2009:312–315. Available at: http://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_0312.pdf (accessed April 7, 2017).

[8] Kalyan, V. P. 2011. Razrabotka algoritmov raspoznavaniya emotsional'nogo sostoyaniya cheloveka po paralingvisticheskim osobennostyam rechi [Development of algorithms for recognizing a person's emotional state by paralinguistic features of speech]. *Dokl. 15-y Vseross. konf. "Matematicheskie metody raspoznavaniya obrazov"* [15th All-Russian Conference on Mathematical Methods of Sspeech Recognition Proceedings] Moscow: MAKS-Press. 334–349.

[9] Brester, K. J. 2013. Kollektivnyy evolyutsionnyy metod mnogokriterial'noy optimizatsii v zadachakh analiza rechevykh signalov [Collective evolutionary method of multicriteria optimization in problems of analysis of speech signals]. PhD Diss. Krasnoyarsk. 143 p. Available at: http://research.sfu-kras.ru/sites/research.sfu-kras.ru/files/Dissertaciya_Brester_K.Yu_.pdf (accessed April 7, 2017).

[10] Kalyan, V. P. 2013. Arkhitektura sistemy raspoznavaniya emotsional'nogo sostoyaniya cheloveka po rechi [Architecture of the system of recognition of the emotional state of a person by speech]. *Modeli i metody raspoznavaniya rechi* [Models and methods of speech recognition]. Moscow: A. A. Dorodnitsyn CC RAS. 89–98.

[11] Kalyan, V. P. 2012. Morfologiya situatsii v systeme raspoznavaniya emotsional'nogo sostoyaniya cheloveka po rechi [Morphology of the situation in the system of recognition of the emotional state of a person by speech]. *Modeli i metody raspoznavaniya rechi* [Models and methods of speech recognition]. Moscow: A. A. Dorodnitsyn CC RAS. 92–102.

[12] Sidorov, K. V., and N. N. Filatova. 2012. Analiz priznakov emotsional'no okrashennoy rechi. *Izvestija UFU* 134(9):39–45. Available at: http://eprints.tstu.tver.ru/69/1/3.pdf (accessed April 7, 2017).

Received September 13, 2016