

Антропоморфный метод распознавания эмоций в звучащей речи

А.В. Вартанов МГУ имени М.В.Ломоносова, Москва, Россия

Поступила: 18 февраля 2013 / Принята к публикации: 3 марта 2013

Anthropomorphic method of emotion recognition in sounding speech

A.V.Vartanov Moscow Lomonosov State University, Moscow, Russia

Received: February 18, 2013 / Accepted for publication: March 3, 2013

Предложен новый эффективный метод автоматического распознавания эмоций по речевому сигналу, основанный на четырехмерной сферической модели эмоций и принципах кодирования информации в нервной системе. В результате разработан и экспериментально протестирован принцип относительного кросс-частотного амплитудно-вариабельного кодирования эмоций в речевом сигнале. Проверялась гипотеза о том, что речь является многоканальным (разнесенным по частотам) сигналом, в каждой полосе которого возможны независимые быстрые микро-изменения амплитуды. Показано соответствие выделенных параметров речевого сигнала и субъективного восприятия тех же образцов (коротких слов «да» и «нет») в системе формализованных параметров четырехмерной психофизиологической модели эмоций. Полученные параметры (факторы) можно охарактеризовать как бимодальные спектральные фильтры. Фактор 1 определяет изменение звукового сигнала по оси знака эмоций — чем больше вклад данного компонента по сравнению с другими, тем положительнее (лучше, полезнее) оценивается объект высказывания. Фактор 2 показывает степень информационной неопределенности ситуации — удивление в противоположность уверенности (спокойствию). Фактор 3 характеризует притяжение (любовь), при этом для набора слов «нет» он сопровождается отсутствием активного отвержения, а для набора «да» — положительной оценкой (знаком). Фактор 4 соответствует характеру отвержения, определяет, будет ли агрессивная (активная) или пассивная (страх, бегство) реакция. Полученные результаты в целом подтверждают продуктивность предлагаемого антропоморфного подхода к разработке технических систем, в частности, к методам обработки речевого сигнала и представления данных. Обнаруженное совпадение подтверждает и выделенные ранее параметры психофизиологической модели, дополнительно обосновывая предпочтительность (по сравнению с другими известными в литературе) именно такой системы классификации эмоций, как с точки зрения размерности, так и в отношении ориентации осей пространства модели.

Ключевые слова: эмоции, речевой сигнал, антропоморфный метод

A new efficient method for automatic emotion recognition by speech signal based on the four-dimensional emotions spherical model and principles of information encoding in the nervous system, is described. As a result, the principle of the relative cross-frequency amplitude-variable encoding of emotions in speech signal is proposed and experimentally tested. The hypothesis on the speech being a multichannel signal (a frequency diversity) with each band having possible independent fast micro amplitude change was tested. The agreement between the selected parameters of the speech signal and the subjective perception of the same samples (short words «yes» and «no») in the system of formalized parameters of psychophysiological emotion for the four-dimensional model is shown. The obtained parameters (factors) may be characterized as bimodal spectral filters. Factor 1 has a basic value is 3000 Hz and the secondary value is 500 Hz. It determines the change in the sound signal in accordance with «character emotion» axis, and the contribution of this component as compared with other components, the more positive (better useful) ones are estimated in the utterance. Factor 2 has two extremes at frequencies that lie somewhere near 1000 and 1750 Hz. It determines the degree of information uncertainty as opposed to confidence (calm). Factor 3 characterizes affection (love). It corresponds to the most widely spaced peaks: low frequencies of about 150 Hz and high-frequencies of 3500 Hz. In yes-no dichotomy «no» is accompanied by the absence of active rejection, and «yes» is characterised as a positive assessment. Factor 4 has similar range between 600 Hz and 1500 Hz. The configuration is close to factor 2, but it is shifted with respect to the low-frequency region, getting their peaks in its local minimum. This component determines whether aggressive (active) or passive (fear, escape) reaction is provoked in the subject. The results obtained confirm the efficiency of the proposed general anthropomorphic approach to the development of technical systems, in particular, the methods of speech signal processing and data presentation. It also confirms the identity of a previously identified psychophysiological model parameters, further justifying the preference (compared with other well-known ones) of this emotions classification, both in terms of dimensions and orientation of the axes of the model space.

Keywords: emotions, speech signal, anthropomorphic approach

Известно, что речь человека, находящегося в различных эмоциональных состояниях, различается по целому ряду показателей. К числу наиболее информативных относят, прежде всего, характеристики просодической группы, которые тонко отражают процессуальную сторону устных высказываний и, в первую очередь, изменяются при реакциях аффективного плана (Златоустова, 1957; Михайлов, Златоустова, 1987; Никишкян, 1987). Задача автоматиче-

ского распознавания звучащей речи и, в частности, ее эмоциональной окрашенности является междисциплинарной и постоянно привлекает исследователей разных специальностей — не только лингвистов, но и математиков, программистов, психологов, физиологов. От ее решения зависит прогресс современных автоматизированных систем управления, реабилитации и протезирования, систем безопасности, срочного оповещения и т.п. Решение этой

задачи имеет большое научное значение для всех сфер фундаментальных исследований человека и информационных технологий. В последние годы явно усилился интерес к анализу речевого сигнала, рассматриваемого в качестве наиболее удобного объективного показателя выражения эмоций, эмоционального состояния человека (Сидоров, Филатова, 2012). Это касается не только сфер деятельности с повышенной ответственностью — космонавтики, авиации (летчики,

диспетчеры аэропорта), обслуживания АЭС и пр., которые изначально доминировали в этом отношении (Хроматиди, 2005; Соловьева, 2008; Chen, 2008; Siging, 2009; Фролов, Милованова, 2009; Розалиев, 2009; Калужный, 2009; Перервенко, 2009; Morist, 2010), но широкой бытовой сферы.

Разработчики новых методов и инструментов анализа пользуются лишь собственным «здравым смыслом» и некоторыми теоретическими обобщениями психологов и фонологов. А последним для анализа эмоциональных явлений приходится пользоваться «стандартными», общедоступными инструментами объективного анализа речевых сигналов.

В интернете, новостных лентах и популярных изданиях периодически появляются сообщения о все более успешных попытках создания программ и бытовых устройств, реагирующих на эмоции в голосе человека. Например, «Ноосфера» сообщает, что «инженеры из Рочестерского университета (Великобритания) разработали программу, способную распознавать эмоции человека по его речи, даже не понимая смысла сказанного. Программа ориентируется на базу звукозаписей, состоящую из календарных дат, произнесенных профессиональными актерами с разными интонациями. Алгоритм анализирует 12 характерных параметров речи, таких как высота и громкость звука. На их основании он определяет одну из шести эмоций. По словам разработчиков, точность распознавания составляет 81 процент — значительно лучше 55 процентов, которых удавалось добиться в предыдущих аналогичных исследованиях. Авторы уже разработали первое коммерческое приложение — программу, отображающую на экране веселый или грустный смай-

лик в зависимости от результата анализа записанного голоса. Это лишь первый этап. Авторы программы фантазируют, что в дальнейшем смартфоны смогут менять цветовую схему интерфейса или выбирать подходящую музыку в зависимости от настроения владельца» (Шпикуляк, 2012). На сайте Animal language

отмечается, что, хотя изучение языка эмоций точными научными методами еще лишь начинается, но уже сейчас стало вырисовываться большое значение этой проблемы, как для теоретической науки, так и для практики (Animal language, 2013). При этом понятно, что решить эту задачу нельзя без знания алфавита акустического языка эмоций. Но, «чтобы заложить этот алфавит в электронный мозг робота, необходимо формализовать признаки, ответственные за эмоциональность голоса» (Animal language, 2013).

Однако, несмотря на множество исследований и коммерческих предложений в данной области, проблема автоматического распознавания эмоционального состояния говорящего по речи на данный момент не является полностью решенной, в частности, отсутствует модель описания речевых образцов в условиях проявления разных видов эмоций (Сидоров, Филатова, 2012). Процесс интерпретации (распознавания) эмоций человека по естественной речи является весьма сложной задачей, как

в области математической формализации задачи, так и в плане поиска способов четкой конкретизации эмоционального состояния — однозначного детектирования эмоции по речевому сигналу. В настоящее время отсутствует универсальная теоретическая модель описания речевых образцов в условиях проявления разных видов эмоций (Филатова, Сидоров, 2012).

Это обусловлено целым комплексом взаимосвязанных проблем. С одной стороны, необходимо выделить в речевом сигнале те параметры, которые могли бы служить индикаторами эмоций. Здесь возникают проблемы их регистрации, математического анализа, поиска соответствующих алгоритмов и технических средств. Для решения этой задачи требуется четко задать «входные» и «выходные» данные, формально представить требуемый результат. С другой стороны, необходимы формальные, объективные методы для систематизации и классификации таких сложных явлений как эмоции человека. Нужно разработать адекватную модель и собрать базу данных — набор соответствующих «образцов» состояний и корреспондирующих им фрагментов речи. Получается порочный круг: чтобы решить одну задачу, надо уже иметь решение другой.

Тем не менее, научные исследования и практические разработки в этом направлении предпринимаются со все большей интенсивностью, подстегиваемой коммерческими возможностями. При этом, как правило, разработчики новых методов и инструментов анализа пользуются лишь собственным «здравым смыслом» и некоторыми теоретическими обобщениями психологов и фонологов. А последним для анализа эмоциональных явлений приходится пользоваться «стандартными», общедоступными инструментами объективного анализа речевых сигналов. Чтобы хоть как-то приблизиться к достижению практической эффективности, всем приходится упрощать задачу — при разработке новых методов анализа речевого сигнала ограничиваться отдельными аспектами эмоциональных феноменов, например, только интерпретацией знака эмоций (Филатова, Сидоров, 2012) или отдельных эмоций, наиболее важных для данной области применения. В итоге общая эффективность предлагаемых



Александр Валентинович Вартанов — кандидат психол. наук, старший научный сотрудник кафедры психофизиологии факультета психологии МГУ имени М.В. Ломоносова.

Развивает авторскую концепцию мозговых механизмов знакового опосредования высших психических функций человека, природы и мозговых механизмов сознания, порождения и понимания речи. Исследует мозговые механизмы когнитивных процессов (восприятие, память, воображение, понимание речи) и эмоций (их проявлений и звучащей речи и семантике). Имеет более 100 научных публикаций.

E-mail: a_v_vartanov@mail.ru

в настоящий момент средств невысока. Приведенный выше пример из «Ноосферы» наглядно это подтверждает: даже при решении задачи по распознаванию всего шести эмоций, результат сводится к примитивному бинарному действию. А метод в типичном случае базируется на стандартных алгоритмах сопоставления с образцом в расчете на простое количественное увеличение быстродействия и объема памяти (например, за счет «облачных» технологий) и размера «словаря» образцов.

О параметрах речевого сигнала

Литературный обзор, проведенный К.В. Сидоровым и Н.Н. Филатовой, показывает, что на современном этапе можно выделить четыре группы объективных признаков и соответствующих методов, позволяющих различать речевые образцы: спектрально-временные, кепстральные, амплитудно-частотные и признаки на основе нелинейной динамики. (Сидоров, Филатова, 2012). Показано, что, основываясь только на одних простых спектральных характеристиках звукового сигнала, невозможно правильно распознавать и идентифицировать различные эмоции (Сидоров, 2011).

Спектрально-временные признаки отражают своеобразие формы временного ряда, спектра голосовых импульсов у разных лиц и специфику фильтрующих функций их речевых трактов. Они характеризуют особенности речевого потока, связанные с динамикой перестройки артикуляционных органов речи говорящего, и являются интегральными характеристиками речевого потока, демонстрирующими своеобразие взаимосвязи или синхронности движения артикуляторных органов говорящего. Амплитудно-частотные признаки также несут важную информацию. Большинство исследований в качестве наиболее информативных акустических коррелятов эмоциональных и функциональных состояний рассматривают ряд частотных, временных и мощностных характеристик голосового сигнала (Адашинская, Чернов, 2007). Как правило, стенические состояния ведут к возрастанию, а астенические – к понижению показателей основного тона, форманта и интенсивности. Обнаружена взаимосвязь акустических параметров речи, эмоциональных и функциональных со-

стояний, обусловленная индивидуальными особенностями говорящих, что выражается в разнонаправленности изменений ряда временных и мощностных параметров речи (Адашинская, Чернов, 2007). Однако применение этих признаков не позволяет в полной мере использовать их в качестве инструмента идентификации эмоционально окрашенной речи (Сидоров, 2011).

Спектрально-временные признаки отражают своеобразие формы временного ряда, спектра голосовых импульсов у разных лиц и специфику фильтрующих функций их речевых трактов. Они характеризуют особенности речевого потока, связанные с динамикой перестройки артикуляционных органов речи говорящего, и являются интегральными характеристиками речевого потока, демонстрирующими своеобразие взаимосвязи или синхронности движения артикуляторных органов говорящего

В группе спектрально-временных признаков были выделены параметры, инвариантные к действию повышенного уровня сигнала, описывающие статистические характеристики речевого сигнала и основного тона, особенности спектральной структуры (Розалиев, 2009). Группа признаков эмоционально окрашенной речи по кепстральным коэффициентам позволяет отделить сигнал возбуждения от сигнала речевого тракта. Мел-частотные кепстральные коэффициенты широко используются в качестве набора признаков речевого сигнала, поскольку они учитывают психоакусти-

но использовать для диагностических правил и признаков, позволяющих распознать и правильно идентифицировать различные эмоции в эмоционально окрашенном речевом сигнале. Так, К.В. Сидоров и Н.Н. Филатова предложили модель интерпретации знака эмоции по правилу объединения нечетких множеств, характеризующих значения R_{max} – усредненного максимального вектора

реконструкции аттрактора по четырем квадрантам (Филатова, Сидоров, 2012). В работе Р.Ю. Романенко рассмотрена возможность применения вейвлет-анализа речевого сигнала с целью использования в системе распознавания речи (Романенко, 2010). Предлагается также проводить классификацию эмоционально окрашенной речи с использованием метода опорных векторов (Хейдоров, 2008). Как отмечалось около десяти лет назад, аппарат акустического анализа речи уже достаточно развит (Бабин, Мазуренко, Холоденко, 2004). Практически все наиболее часто используемые способы расчета

Речь, порождаемая человеком, находящимся в различных эмоциональных состояниях, характеризуется целым рядом показателей, в том числе таких, которые могут отражать процессуальную сторону устных высказываний. Однако, формальные критерии не могут дать общей картины изменения текущего состояния и отношения человека, поскольку не разработана антропоморфная система классификации эмоциональных проявлений в звучащей речи

ческие принципы восприятия речи и мел-шкалу, связанную с критическими полосами слуха (Siging, 2010; Сидоров, Филатова, 2012). Для группы признаков нелинейной динамики речевой сигнал рассматривается как скалярная величина, наблюдаемая в системе голосового тракта человека (Старченко и др., 2010). В настоящее время методы нелинейной динамики и нелинейной авторегрессии позволяют восстанавливать фазовый портрет аттрактора по временному ряду или по одной его координате. Экспериментально подтверждено, что выявленные отличия в форме аттракторов мож-

акустических параметров речевого сигнала реализованы в известных и общедоступных математических компьютерных пакетах обработки сигналов (Бабин, Мазуренко, Холоденко, 2004), например, в пакетах SPL и IPPS фирмы Intel (Intel Developer Centers, 2013).

Таким образом, речь, порождаемая человеком, находящимся в различных эмоциональных состояниях, характеризуется целым рядом показателей, в том числе таких, которые могут отражать процессуальную сторону устных высказываний. Однако, формальные критерии, хотя и позволяющие успешно диф-

ференцировать отдельные эмоции по речевым образцам, не могут дать общей картины изменения текущего состояния и отношения человека, поскольку не разработана антропоморфная система классификации эмоциональных проявлений в звучащей речи. Отправной точкой решения вышеописанной проблемы должна стать система, достаточно полно моделирующая процесс восприятия эмоций человеком, которая учитывает совокупность разных аспектов их проявления, в том числе, в речи. Многомерность эмоций, их проявление на различ-

в некоторой степени, воспроизводить процессы, позволяющие нервной системе человека правильно распознавать всю гамму эмоций, т.е. необходима антропоморфная модель эмоций.

Четырехмерная сферическая модель эмоций

Несмотря на всю сложность проблемы, предпринятое ранее исследование эмоциональных характеристик звучащего слова и семантики эмоций позволили построить универсальную четырехмерную сферическую модель

Построение модели проводилось экспериментально с помощью многомерного шкалирования субъективных различий между эмоциональными состояниями, задаваемыми специально созданными образцами. Чтобы урвать и сделать определенным содержание этих образцов, в эксперименте использовалось одно и то же слово, произнесенное в разных эмоциональных состояниях

ных уровнях отражения и деятельности, способность к слиянию и образованию сочетаний исключают возможность их простой линейной классификации (Виллюнас, 1984) или создания конечного дискретного набора определенных вариантов. Обычно выделяют как минимум десять типов эмоциональных отношений или так называемых фундаментальных эмоций, между которыми, однако, возможны плавные переходы.

Эти типы в достаточной мере условны, обозначая (в виде понятийных категорий) лишь наиболее важные места эмоционального континуума. Поэтому в разное время на основе различных экспериментальных методов и эмпирических фактов делались попытки выделить в этом разнообразии ограниченное число базовых факторов или основных «компонентов эмоционального качества», которые бы выступали по отношению к отдельным эмоциональным переживаниям как родовые исходные характеристики или «образующие». В настоящее время известен целый ряд таких независимых или частично перекрывающихся признаков и оснований для деления эмоциональных явлений. Это объясняется тем, что эмоции проявляются одновременно и во внутренних переживаниях, и в поведении, причем, и то, и другое обусловлено еще специфической физиологической активацией. При этом аппарат анализа речевого сигнала также должен, хотя бы

эмоций (Виденеева, Хлудова, Вартанов, 2000; Вартанов, Виденеева, 2001; Вартанов, Вартанова, 2003; Вартанов, Вартанова, 2005). Эта модель объективирует и формализует в системе четырех количественных параметров все многообразие переживаний и различные проявления эмоций в речи, мимике, а также в семантике.

Построение модели проводилось экспериментально с помощью многомерного шкалирования субъективных различий между эмоциональными состояниями, задаваемыми специально созданными образцами. Чтобы урвать 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». Уже такие короткие одноударные слова, как свидетельствует практика актерского мастерства (Станиславский, 1959), вполне могут адекватно и полно отражать весь спектр эмоциональных проявлений. Эти слова, по сравнению с другими, несут более определенное и независимое от контекста значение, но, в то же время, они более нейтральны и допускают больше вариантов эмоциональной окраски при их произнесении. Из большого числа образцов, наигранных профессиональными актерами и «подловленных» в естественных условиях, было отобрано для каждого набора по 20 на-

иболее удачных, отражающих 10 типичных эмоций, наиболее существенных для актерского исполнения (Станиславский, 1959). Наличие двух наборов таких образцов (противоположных по семантике) позволяет найти универсальные, независимые от конкретного слова параметры, определяющие именно проявление эмоций в речи.

В эксперименте регистрировались субъективные оценки степени попарного различия между звуковыми стимулами. Набор из 20 образцов в каждой из серий образовывал по 190 вариантов пар. Каждая пара предъявлялась не менее чем по 3 раза, т.е. всего 570 пар, которые следовали в случайном порядке. В экспериментах участвовало в общей сложности 25 взрослых испытуемых и 30 детей разных возрастов (с 1-го по 8-й классы). Кроме того, тем же методом исследовалась и семантика эмоций русского языка, для чего использовались различные наборы слов, обозначающих эмоции. Обнаружено, что и дети, и все взрослые одинаково успешно воспринимают и непосредственно сравнивают эмоциональные состояния другого, выраженные в интонациях речи – полученные матрицы всех испытуемых хорошо совпадали (коррелировали) друг с другом, что позволило далее объединить все данные и уменьшить случайный шум получаемых оценок, образуя матрицу различий.

Анализ (метрическим методом) многомерного шкалирования усредненных матриц различий в соответствующих сериях показал, что размерность полученного эмоционального пространства по всем критериям должна быть оценена как равная четырем. Расположение точек-стимулов в четырехмерном пространстве проверялось на сферичность. Оказалось, что в серии «да» вариативность радиуса четырехмерной сферы составляла всего 9,71%, а в серии «нет» – 9,94%. Это хорошо согласуется с теоретическими разработками о принципах кодирования в нервной системе (Соколов, Вайткявичюс, 1989; Соколов, 2001; Вартанов, 2011), на основе которых может быть построена антропоморфная нейротропная модель эмоций.

После вращения евклидовы оси пространства получили интерпретацию как определенные нейронные (мозговые) механизмы эмоций, а угловые характе-

ристики – как субъективные качества эмоций. Первые две евклидовы оси пространства связаны с оценкой ситуации: ось 1 – по знаку (хорошо, полезно, приятно или плохо, вредно, неприятно), ось 2 – по степени информационной определенности (уверенность – удивление). Система третьей и четвертой осей связана с побуждением: ось 3 – притяжение, ось 4 – отвержение (оборонительная реакция), активное (агрессия) или пассивное (страх, затаивание) избегание. Это хорошо согласуется с известными мозговыми механизмами эмоций (Симонов, 1981; 2001). Так, ось 3 и положительное направление оси 1 (вроде бы сходные качества) отражают работу разных групп нейронов гипоталамуса – побудительных и подкрепляющих, которые хотя и определяют, казалось бы, одни и те же положительные эмоциональные состояния, но находятся между собой в конкурентных отношениях (что проявляется в ортогональности осей модели). Ось 2 и отрицательное направление оси 1 можно связать с работой гиппокампа (активизирующегося в условиях информационной неопределенности) и фронтальной коры (дорсальной ее части), а также с миндалины лимбической системой – вентральной части префронтальной коры. В целом префронтальная кора, являясь, как и гиппокамп, «информационной» структурой мозга, ориентирует поведение на сигналы высоковероятных событий. Ось 4, которая делит активные и пассивные оборонительные реакции, по-видимому, также описывает активность медиального гипоталамуса, точнее двух его структур, стимуляция которых вызывает оборонительные реакции нападения (положительное направление оси 4) или бегства, соответственно (отрицательное направление оси 4).

Оказалось, что три угла четырехмерной гиперболы, выбранные в проекции осей 1-2, 3-4 и угол, образуемый движением точки между двумя этими плоскостями, задают такие субъективно переживаемые качества эмоций, которые описывал еще В. Вундт (Вундт, 1984). Это три качества: 1) эмоциональный тон (удовольствие – неудовольствие), 2) возбуждение – успокоение – угнетение, 3) напряжение – разрешение. При этом первый и второй углы упорядочивают все 10 основных эмоций по модально-

сти: 5 эмоций, определяемых ситуацией и 5, определяемых собственной активностью. Но оказалось также, что при выборе другой системы угловых параметров – если взять три угла в системе осей 4-1, 3-2 и угол, образуемый движением точки между этими плоскостями, обнаруживается другая система классификации эмоций, описываемая при исследовании выражений лица – круговая система Х. Шлосберга ([Schlosberg, 1941) и сферическая модель Ч.А. Измайлова (Измайлов, Коршунова, Соколов, 1999), а также семантика Ч. Осгуд

Предлагаемая четырехмерная сферическая модель может служить общей классификационной системой для эмоциональных явлений, объединяя, как физиологические представления о мозговых механизмах эмоциональной регуляции, так и известные психологические классификации, полученные на основе разных экспериментальных данных

(Osgood, Suci, Tannenbaum, 1957). Она включает: 1) эмоциональный тон или знак (упорядочивает 6 основных эмоций по модальности), 2) активность или яркость эмоций (возбуждение – покой) и 3) эмоциональная насыщенность (сила проявления эмоций).

Таким образом, полученные данные показывают, что звучащая речь вполне определенно и достаточно точно выражает эмоциональное состояние говорящего, хорошо корреспондируя с другими важными для человека каналами: зрительным восприятием (по мимике и выразительным движениям), ощущением своего собственного состояния в самонаблюдении. Она также закреплена в языковых терминах (общественный опыт обозначения эмоций в социальном канале коммуникации). Предлагаемая четырехмерная сферическая модель может служить общей классификационной системой для эмоциональных явлений, объединяя, как физиологические представления о мозговых механизмах эмоциональной регуляции, так и известные психологические классификации, полученные на основе разных экспериментальных данных. Она также количественно объясняет все возможные нюансы и плавные переходы эмоций, представляя каждую конкретную эмоцию как линейную комбинацию выделенных основных психофизиологических параметров. По-видимому, у человека и животных существует специальный механизм эмоционального

или чувственного отражения, необходимый для регуляции поведения и ориентировки в ситуации, работа которого может быть формально представлена в виде вышеописанной четырехмерной сферической модели. Наличие единого механизма во всех процессах позволяет представить все эмоциональные явления в одной и той же системе параметров. В результате данная модель, являясь антропоморфной (поскольку отражает субъективное отношение человека) и нейротропной (поскольку отражает нейронные механизмы), позво-

ляет количественно описать и наглядно представить изменения текущего состояния человека или его эмоционального отношения. Она может стать базисом при конструировании устройства, которое в удобной форме представляет детектируемые по звучащей речи эмоциональные состояния человека.

Результаты выявления параметров речевого сигнала в соответствии с предлагаемой антропоморфной моделью

В качестве исходного материала для выявления параметров речевого сигнала, которые должны воспроизводить параметры вышеописанной сферической модели эмоций, были использованы те же образцы звуковых фрагментов, что и в эксперименте с субъективными оценками. Это – 20 образцов слова «да» и 20 образцов слова «нет» (средняя длительность 0,60 сек, стандартное отклонение 0,19 сек; минимальная длительность 0,3 сек, максимальная 0,98 сек; запись в полосе до 8000 Гц). После исследования возможных параметров, наиболее полно представляющих свойства данного набора образцов, было обнаружено, что наилучшим образом поставленной задаче соответствует показатель, вычисляемый по следующему алгоритму:

1 Для звукового фрагмента с помощью стандартных средств – быстрое преобразование Фурье со сглаживанием в минимальном скользящем окне порядка 10-15 мс вычисляется последо-

вательностью мгновенных спектров мощности сигнала (в диапазон от 0 до 4000 Гц с шагом 50 Гц).

2 На основе последовательности мгновенных спектров в скользящем окне (исследовались окна порядка 50-200 мс) вычисляется показатель микро-вариативности (стандартное отклонение) амплитуды (квадратного корня от мощности) на каждой частоте.

3 Для вычисления интегральной оценки всего звукового образца использовалось простое усреднение предыдущего показателя по всему интервалу звучания и получения одного вектора (по частоте) для каждого звукового образца.

Такой алгоритм был выбран на основе теоретических предположений об общих принципах кодирования информации в нервной системе (Варганов, 2011). Дополнительным основанием послужили наблюдения, впервые сделанные еще Ч. Дарвином о том, что эмоциональную выразительность голоса придает именно определенное «дрожание» тембра, что особенно важно для выразительности пения (Дарвин, 1940). Как отмечалось многими авторами, изменения громкости речи в макро-варианте на протяжении всего высказывания также может характеризовать эмоциональное отношение говорящего. Однако и быстрые микро-изменения амплитуды (в пределах короткого слова или междоуметия) также могут служить мерой изменения эмоционального состояния или отношения человека. При этом, для того, чтобы было возможно передать всю гамму эмоций, как показано выше, недостаточно только одного параметра, поэтому проверялась гипотеза о том, что речь – это многоканальный (разнесенный по частотам) сигнал, в каждой полосе которого возможны независимые быстрые микро-изменения амплитуды. То есть, основное предположение свелось к проверке относительного кросс частотного амплитудно-вариативного кодирования эмоций в речевом сигнале.

Все полученные звуковые образцы (40 записей разной длины) были обработаны с помощью специально созданных программных средств, а усредненные значения предлагаемого параметра в исследованном частотном диапазоне (с шагом 50 Гц) были собраны в единый массив данных, который далее подвергся статистическому (факторному)

анализу. Вращение и интерпретация полученных факторов проводились с помощью специально разработанных средств на основе сопоставления с известными для данных образцов (наборов слов «да» и «нет») оценками в четырехмерной модели эмоций.

В результате факторный анализ позволил оценить размерность и выявить 4 фактора (рис. 1), которые совокупно описывают 70,15% всей дисперсии данных.

После специального вращения в пространстве модели эмоций для достижения наилучшего соответствия между

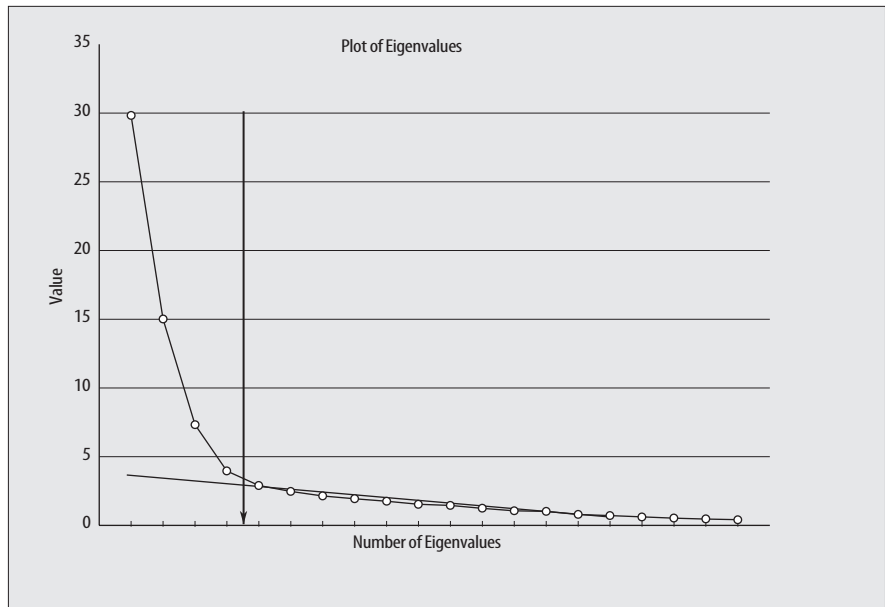


Рис. 1. График распределения собственных значений при факторном анализе всего набора звуковых образцов, включая слова «да» и «нет» (всего 40 образцов). Стрелками отмечена граница, в соответствии с которой можно оценить размерность факторного пространства как равную четырем.

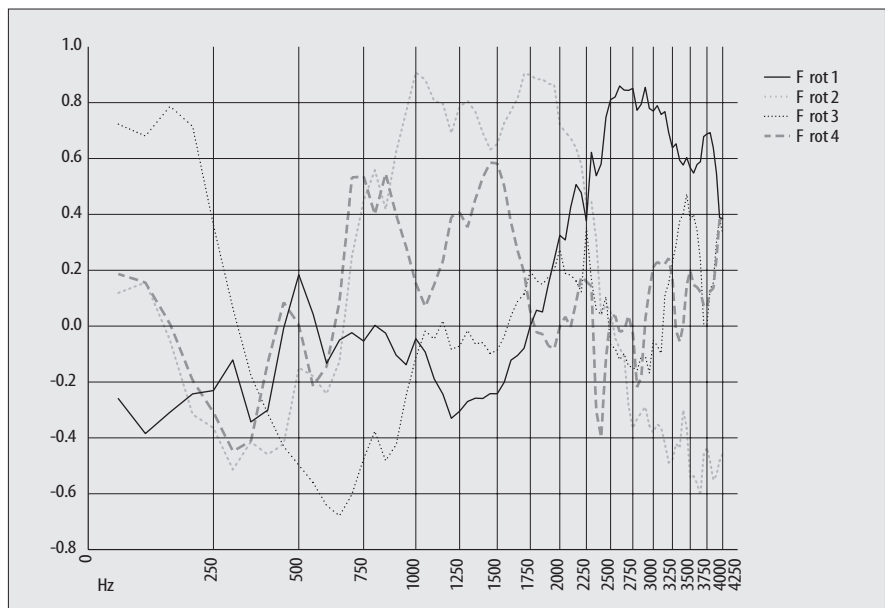


Рис. 2. Спектральные характеристики четырех выделенных факторов, которые описывают эмоциональные качества всех речевых образцов в совокупности. Горизонтальная ось дана в логарифмическом масштабе. В результате вычисления значения этих факторов и их нормализации (как этого требует теория кодирования и сферичность пространства психофизиологической модели) было проведено сопоставление оценок, полученных путем формального анализа звукового сигнала и субъективных оценок, в соответствии с моделью эмоций. Вычисленные коэффициенты корреляции для каждого набора в отдельности (слова «да» и «нет») и совместно представлены в таблицах 1-3.

нормированными значениями факторов и координатами образцов факторы получили спектральное выражение, показанное на рис. 2. Решение, полученное таким методом вращения, не сильно отличалось от решения, полученного методом варимакс-нормализацией. В результате полученные факторы можно охарактеризовать как бимодальные спектральные фильтры. Фактор 1 имеет основной максимум в области 3000 Гц и вспомогательный – 500 Гц. Фактор 2 имеет два близких максимума на частотах 1000 и 1750 Гц. Фактор 3 имеет самые широко разнесенные максимумы – в низкочастотной области (около 150 Гц) и высокочастотной области (3500 Гц). Фактор 4 имеет близкие максимумы на 600 и 1500 Гц и близок к фактору 2, но сдвинут относительно него в низкочастотную область, попадая своими максимумами в его локальные минимумы.

	x1	x2	x3	x4
Factor 1	0.42	-0.36	0.13	-0.08
Factor 2	-0.36	0.59	0.11	-0.17
Factor 3	0.11	0.11	0.65	-0.30
Factor 4	-0.10	-0.15	-0.27	0.63

Таблица 1. Коэффициенты корреляции Пирсона между параметрами модели эмоций (x1-x4) и параметрами речевого сигнала (факторами). Жирным курсивом показаны значимые ($p < .05$ при $N=40$) коэффициенты.

	x1	x2	x3	x4
Factor 1	0.57	-0.24	0.35	-0.22
Factor 2	-0.14	0.28	0.78	-0.51
Factor 3	0.53	0.07	0.50	0.04
Factor 4	-0.03	-0.41	-0.57	0.72

Таблица 2. Коэффициенты корреляции Пирсона для набора «да» между параметрами модели эмоций (x1-x4) и параметрами речевого сигнала (факторами). Жирным курсивом показаны значимые ($p < .05$ при $N=20$) коэффициенты.

	x1	x2	x3	x4
Factor 1	0.15	-0.46	-0.23	0.15
Factor 2	-0.50	0.55	-0.14	-0.43
Factor 3	-0.31	0.41	0.79	-0.58
Factor 4	-0.05	-0.22	0.12	0.55

Таблица 3. Коэффициенты корреляции Пирсона для набора «нет» между параметрами модели эмоций (x1-x4) и параметрами речевого сигнала (факторами). Жирным курсивом показаны значимые ($p < .05$ при $N=20$) коэффициенты.

При анализе этих данных необходимо учесть, что, хотя и выделенные факторы и параметры (оси) модели эмоций ортогональны, тем не менее, исследуемые образцы не заполняют все пространство равномерно и существенно различаются для наборов «да» и «нет».

Для того, чтобы было возможно передать всю гамму эмоций недостаточно только одного параметра, поэтому проверялась гипотеза о том, что речь – это многоканальный (разнесенный по частотам) сигнал, в каждой полосе которого возможны независимые быстрые микро-изменения амплитуды. То есть, основное предположение свелось к проверке относительного кросс-частотного амплитудно-вариабельного кодирования эмоций

Поэтому сами координаты образцов в модели в некоторой степени коррелируют между собой (таблица 4). Похожая корреляция по той же причине наблюдается и между значениями факторов, что понятно, если система голосовых параметров (выделенных факторов) и система психофизиологических параметров модели близки.

	x1	x2	x3	x4
x1	1.00	-0.22	0.07	-0.03
x2	-0.22	1.00	0.34	-0.30
x3	0.07	0.34	1.00	-0.44
x4	-0.03	-0.30	-0.44	1.00

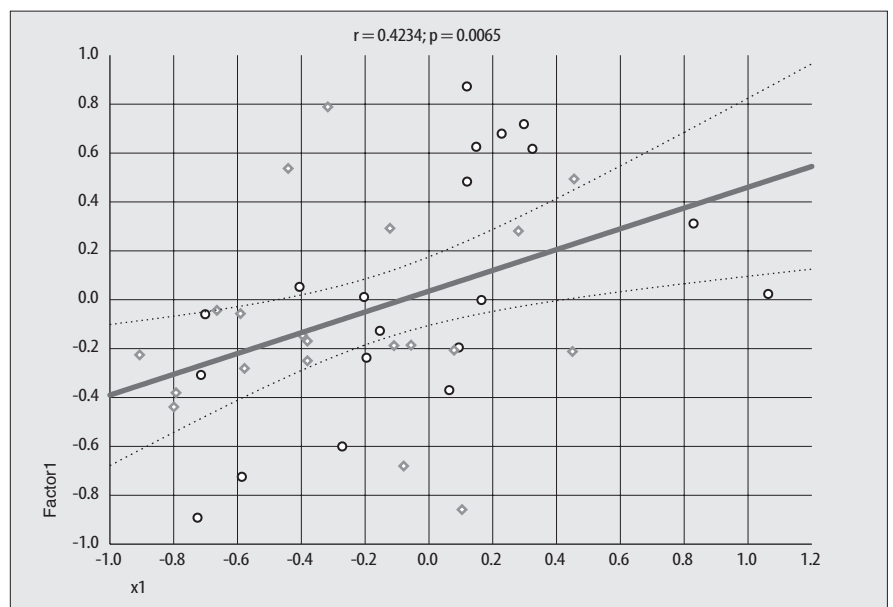
Таблица 4. Коэффициенты корреляции Пирсона между параметрами модели эмоций (x1-x4). Жирным курсивом показаны значимые ($p < .05$ при $N=40$) коэффициенты.

Корреляционные поля для выделенных факторов представлены на рис. 3.

На основании этих данных можно заключить, что в целом первые четыре спектральных параметра значимо коррелируют с параметрами психофизиологической модели. При этом наблюдается определенное своеобразие связей в зависимости от набора образцов, что

позволяет предположить, что семантическое значение слова («да» или «нет») в некоторой степени определяет и направление изменения данных параметров голоса. Тем не менее, можно заключить, что первый фактор определяет изменение звукового сигнала по оси знака эмоций – чем больше вклад данного компонента по сравнению с другими, тем положительнее (лучше, полезнее) оценивается объект высказывания.

Это более справедливо для утверждений в наборе «да». Второй спектральный параметр в целом и в наборе «нет» определяет степень информационной неопределенности ситуации – удивление в противоположность уверенности (спокойствию). При этом для слов «да» это удивление сопровождается также еще влечением, а «не отвержением», т.е. характеризует любопытство в случае согласия или чистое удивление в случае отрицания. Третий компонент в целом



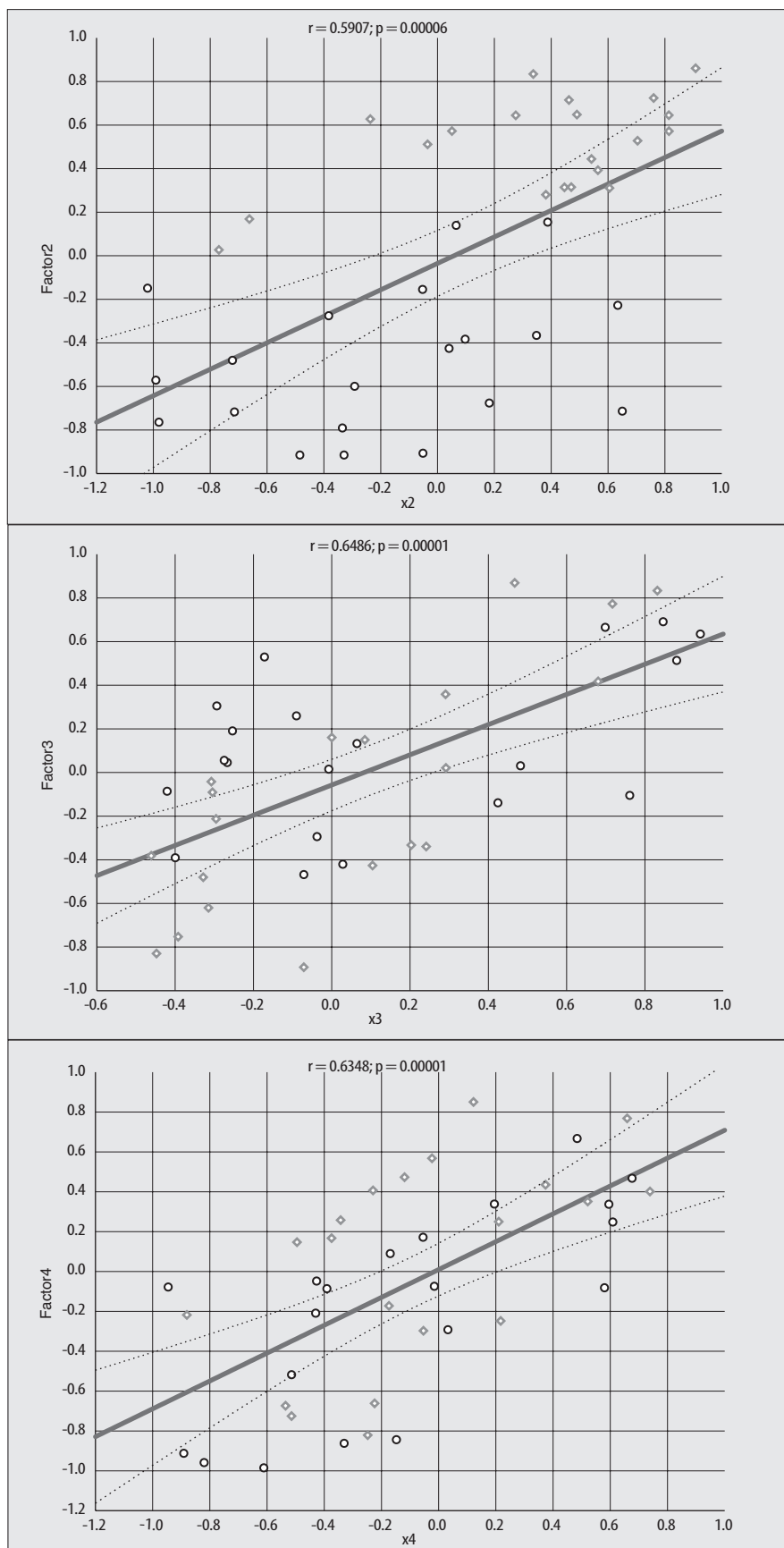


Рис. 3. Корреляционные поля соответствия выделенных параметров речевого сигнала и психофизиологической модели. Кругами обозначены образцы «да», а квадратами – «нет».

и во всех наборах в отдельности характеризует притяжение (любовь), при этом, для набора «нет» он сопровождается отсутствием активного отвержения, а для набора «да» – положительной оценкой (знаком). Четвертый компонент соответствует, как в целом, так и для обоих наборов по отдельности, степени и характеру отвержения. Он определяет: будет ли реакция агрессивная (активная) или пассивная (страх, бегство). При этом в наборе «да» он характеризуется еще «не притяжением».

Заключение

Проведенный анализ и полученные в результате параметры звукового сигнала позволяют построить эффективный антропоморфный (и по процессу, и по результату) метод диагностики и представления эмоций в звучащей речи. Хорошее согласие параметров речевого сигнала и психофизиологической модели эмоций подтверждают теоретические представления о принципах кодирования информации в нервной системе и продуктивности предлагаемого антропоморфного подхода к разработке технических систем, в частности методов обработки речевого сигнала.

Обнаруженное совпадение подтверждает и выделенные ранее параметры психофизиологической модели, дополнительно обосновывая предпочтительность (по сравнению с другими известными в литературе) именно такой системы классификации эмоций, как с точки зрения размерности, так и в отношении ориентации осей пространства модели.

Полученные данные также ставят новые вопросы о взаимосвязи и взаимодействии разноуровневых систем управления – вербальной сознательной и эмоциональной досознательной, которые совместно отражаются в речевом сигнале. Необходимо также провести дополнительное исследование универсальности выделенных параметров речевого сигнала по отношению к специфике голоса диктора (в данном исследовании описан голос только одного диктора) и различные речевые высказывания, поскольку возможна интерференция содержания и просодического оформления высказывания, а также интерференция параметров, кодирующих эмоциональное и вербальное содержание речевого сигнала.

Литература:

- Адашинская Г.А. Акустические корреляты индивидуальных особенностей функциональных и эмоциональных состояний / Г.А. Адашинская, Д.Н. Чернов // *Авиакосмическая и экологическая медицина*. – 2007. – Т. 41. – № 2. – С. 3-13.
- Бабин Д.Н. О перспективах создания системы автоматического распознавания слитной устной русской речи / Д.Н. Бабин, И.Л. Мазуренко, А.Б. Холоденко // *Интеллектуальные системы*. 2004. – Т. 8. – Вып. 1-4. – С. 45-70.
- Вартанов А.В. Механизмы семантики: человек – нейрон – модель / А.В. Вартанов // *Нейрокомпьютеры: разработка, применение*. – 2011. – № 12. – С. 54-64.
- Вартанов А.В. Что такое эмоции? 4-х мерная сферическая модель аспектов переживания, выражения, восприятия и обозначения эмоций / А.В. Вартанов, И.И. Вартанова // *Культурно исторический подход и проблема творчества: Материалы вторых чтений памяти Л.С. Выготского* : сб. / под ред. Е.Е. Кравцовой, В.Ф. Спиридонова, Ю.Е. Кравченко. – Москва : РГГУ, фонд им. Л.С. Выготского, 2003. – С. 13-29.
- Вартанов А.В. Эмоции, мотивация, потребность в филогенезе психики и мозга / А.В. Вартанов, И.И. Вартанова // *Вестник Московского Университета. Сер. 14 Психология*. – 2005. – № 3. – С. 20-35.
- Вартанов А.В. Четырехмерная сферическая модель эмоций и дистанционный речевой контроль состояния человека / А.В. Вартанов, Н.М. Виденеева // *Тезисы докладов рабочей группы «Влияние информационных технологий на национальную безопасность»*. 4-я Ежегодная Конференция Консорциума ПрМ «Построение стратегического сообщества через образование и науку». – Москва, 2001. – 35 с.
- Виденеева Н.М. Эмоциональные характеристики звучащего слова / Н.М. Виденеева, О.О. Хлудова, А.В. Вартанов // *Журнал высшей нервной деятельности*. – 2000. – Т. 50. – Вып. 1. – С. 29-43.
- Виллюнас В.К. Основные проблемы психологической теории эмоций / В.К. Виллюнас // *Психология эмоций : тексты* / под ред. В.К. Виллюнаса, Ю.Б. Гиппенрейтер. – Москва : Изд-во МГУ, 1984. – С. 3-26.
- Вундт В. Психология душевных волнений / В. Вундт // *Психология эмоций : тексты* / под ред. В.К. Виллюнаса, Ю.Б. Гиппенрейтер. – Москва : Изд-во МГУ, 1984. – С. 48-63.
- Дарвин Ч. Выражение эмоций у человека и животных // Дарвин Ч. Сочинения / под ред. Н.П. Павловского. Т. 5. – Москва, Ленинград, 1940.
- Златоустова Л.В. Типы эмфатического ударения в русском литературном языке / Л.В. Златоустова // *Общеуниверситетский сборник*. – 1957. – Т. 117. – С. 107-111.
- Измайлов Ч.А. Сферическая модель различения эмоциональных выражений схематического лица человека / Ч.А. Измайлов, С.Г. Коршунова, Е.Н. Соколов // *Журнал высшей нервной деятельности*. – 1999. – Т. 49. – Вып. 2. – С. 186-199.
- Калужный М.В. Система реабилитации слабовидящих на основе настраиваемой сегментарной модели синтезируемой речи : дис. ...канд. тех. наук / М.В. Калужный. – Санкт-Петербург, 2009.
- Михайлов В.Т., Златоустова Л.В. Измерения параметров речи / В.Т. Михайлов, Л.В. Златоустова. – Москва : Радио и связь, 1987 – 167,[1] с. : ил.
- Никишкян Э.А. Типология интонации эмоциональной речи / Э.А. Никишкян. – Киев, Одесса, 1986.
- Перервенко, Ю.С. Исследование инвариантов нелинейной динамики речи и принципы построения системы аудиоанализа психофизиологического состояния: дис. ...канд. тех. наук / Ю.С. Перервенко. – Таганрог, 2009.
- Розалиев В.Л. Моделирование эмоциональных реакций пользователя при речевом взаимодействии с автоматизированной системой : дис. ... канд. тех. наук / В.Л. Розалиев. – Волгоград: ВГТУ, 2009.
- Романенко Р.Ю. Вейвлет-анализ речевых сигналов. Успехи современной радиоэлектроники / Р.В. Романенко // *Зарубежная радиоэлектроника*. – 2010. – № 12. – С. 51-54.
- Сидоров К.В. Анализ признаков эмоционально окрашенной речи / К.В. Сидоров, Н.Н. Филатов // *Вестник Тверского государственного технического университета*. – 2012. – Вып. 20. – С. 26-31.
- Сидоров К.В. К вопросу оценки эмоциональности естественной и синтезированной речи по объективным признакам / К.В. Сидоров, М.В. Калужный // *Вестник Тверского государственного технического университета*. – 2011. – Вып. 18. – С. 81-85.
- Симонов П.В. Лекции о работе головного мозга: потребностно-информационная теория высшей нервной деятельности / П.В. Симонов. – Москва : Наука, 2001. – 95 с.
- Симонов П.В. Эмоциональный мозг / П.В. Симонов. – Москва : Наука, 1981.
- Соколов Е.Н. Сферическая модель интеллектуальных операций / Е.Н. Соколов // *Психологический журнал*. – 2001. – Т. 22. – № 3. – С. 49-56.
- Соколов Е.Н., Вайтнявичюс Г.Г. Нейроинтеллект: от нейрона к нейрокомпьютеру / Е.Н. Соколов, Г.Г. Вайтнявичюс. – Москва : Наука, 1989. – 238 с.
- Соловьева Е.С. Методы и алгоритмы обработки, анализа речевого сигнала для решения задач голосовой биометрии : дис. ...канд. тех. наук / Е.С. Соловьева. – Москва, 2008.
- Станиславский К.С. Моя жизнь в искусстве / К.С. Станиславский. – Москва, 1959.
- Старченко И.Б. Методы нелинейной динамики для биомедицинских приложений / И.Б. Старченко, Ю.С. Перервенко, О.С. Борисова, Т.В. Момот // *Известия ЮФУ. Технические науки*. – 2010. – № 9 (110). – С. 42-51.
- Филатова Н.Н., Сидоров К.В. Модель интерпретации знака эмоций по естественной речи / Н.Н. Филатова, К.В. Сидоров // *Известия ЮФУ. Технические науки Тематический выпуск*. – 2012. – Т. 134. – № 9 – С. 39-45.
- Фролов М.В., Милованова Г.Б. Речевой сигнал как показатель функционального состояния человека-оператора / М.В. Фролов, Г.Б. Милованова // *Биомедицинская радиоэлектроника*. – 2009. – № 6 – С. 49-53.
- Шпикуляк И. Ему не все равно: смартфоны смогут различать эмоции. // *Ноосфера, IT и электроника* – [Электронный ресурс]. – Режим доступа : <http://noos.com.ua/ru/post/3104/> – Дата обращения 06 декабря 2012.
- Хейдоров И.Э. Классификация эмоционально окрашенной речи с использованием метода опорных векторов / И.Э. Хейдоров, Я. Цзинбинь и др. // *Речевые технологии*. – 2008. – Вып. 3. – С. 63–71.
- Хмылёв В.Л., Кондрасюк В.А. Практические аспекты взаимодействия информационных уровней восприятия реальности // *Вестник Томского*

гос. университета. – 2014. – № 381. – С. 87-93.

Хроматиди А.Ф. Исследование психофизиологического состояния человека на основе эмоциональных признаков речи : дис. ... канд. тех. наук / А.Ф. Хроматиди. – Таганрог, 2005.

Animal language – [Электронный ресурс]. – Режим доступа : http://animalang.biggo.ru/prakticheskoe_znachenie_ – Дата обращения 12.02.2013.

Chen Y.T. A study of emotion recognition on mandarin speech and its performance evaluation : Ph. D. dissertation / Y.T. Chen. – Tatung, 2008.

Intel Developer Centers – [Электронный ресурс]. – Режим доступа : <http://developer.intel.com>; <http://www.intel.com/content/www/us/en/search.html?keyword=SPL+> <http://www.intel.com/content/www/us/en/search.html?context=767188&tab=767189&keyword=IPPS> – Дата обращения 12.12.2013

Morist M.U. Emotional speech synthesis for a radio dj: corpus design and expression modeling: master thesis MTG-UPF dissertation / M.U. Morist. – Barcelona, 2010.

Osgood C.E., Suci G.J. & Tannenbaum P.H. The measurement of meaning. – Urbana. University of Illinois Press. 1957.

Schlosberg H.S. A scale for the judgement of facial expressions // Experimental Psychology. – 1941. – P. 497-510.

Siging W. Recognition of human emotion in speech using modulation spectral features and support vector machines: master of science dissertation / W. Singing. – Kingston, 2009.

References:

Adashinskaja, G.A. (2007) Akusticheskie korreljaty individual'nyh osobennostej funkcional'nyh i jemocional'nyh sostojanij [Acoustic correlates of individual features of functional and emotional states]. Adashinskaja, G.A., & Chernov, D.N., Aviakosmicheskaja i jekologicheskaja medicina [Aerospace and Environmental Medicine]. Vol. 41, 2, pp. 3-13.

Animal language - [electronic resource]. - Mode of access: http://animalang.biggo.ru/prakticheskoe_znachenie_ - Access Date 12.02.2013 .

Babin, D.N. (2004) O perspektivah sozdaniya sistemy avtomaticheskogo raspoznavaniya slitnoj ustnoj russkoj rechi [On the prospects of establishing a system of automatic recognition of continuous speech spoken Russian]. Babin, D.N., Mazurenko, I.L., & Holodenko, A.B. Intellekual'nye sistemy [Intelligent Systems]. Vol. 8, 1-4, pp. 45-70.

Chen, Y.T. (2008) A study of emotion recognition on mandarin speech and its performance evaluation: Ph. D. Thesis / Y.T. Chen. Tatung. 3. Darwin, Ch. (1940) Vyrazhenie emotsij u cheloveka i zhivotnykh [Expression of the Emotions in Man and Animals].

Darvin, Ch. (1940) Sochinenija [Oeuvre] ed. Pavlovskiy, N.P. Vol. 5. Moscow, Leningrad.

Filatova, N.N., & Sidorov, K.V. (2012) Model' interpretacii znaka jemocij po estestvennoj rechi [Model to interpret the sign of emotion on natural speech]. Filatova, N.N., & Sidorov, K.V. Izvestija JuFU. Tehnicheskie nauki Tematicheskij vypusk [Proceedings of the SFU. Engineering Special Issue]. Vol. 134, 9, pp. 39-45.

Frolov, M.V., & Milovanova, G.B. (2009) Rechevoj signal kak pokazatel' funkcional'nogo sostojanija cheloveka-operatora [Speech signal as an indicator of the functional state of the human operator]. Frolov, M.V., & Milovanova, G.B. Biomedicinskaja radioelektronika [Biomedical radio electronics]. 6, pp. 49-53.

Intel Developer Centers - [electronic resource]. - Mode of access : <http://developer.intel.com>; <http://www.intel.com/content/www/us/en/search.html?keyword=SPL+> <http://www.intel.com/content/www/us/en/search.html?context=767188&tab=767189&keyword=IPPS> - access date 12/12/2013

Izmalov, Ch.A. (1999) Sfericheskaja model' razlichenija jemocional'nyh vyrazhenij shematicheskogo lica cheloveka [Spherical model of distinguishing emotional expressions in the human face model]. Izmajlov, Ch.A., Korshunova, S.G., & Sokolov, E.N. Zhurnal vysshej nervnoj dejatel'nosti [Journal of Higher Nervous Activity]. Vol. 49, Issue 2, pp. 186-199.

Kalyuzhnyy, M.V. (2009) Sistema reabilitacii slabovidjashhih na osnove nastraivaemoj segmentarnoj modeli sinteziruemoj rechi: dissertatsiya kandidata tekhnicheskikh nauk [Rehabilitation system for the visually impaired based on custom segmental pattern of synthesized speech, Ph.D. Thesis].

Kalyuzhnyy, M.V. St-Peterburg.

Kheydorov, I.Je. (2008) Klassifikacija jemocional'no okrashennoj rechi s ispol'zovaniem metoda opornykh vektorov [Classification of emotional speech using support vector]. Kheydorov, I.Je., Czinbin' Ja., et al. Rechevyte tehnologii [Speech technologies]. Issue 3, pp. 63-71.

Khromatidi, A.F. (2005) Issledovanie psihofiziologicheskogo sostojanija cheloveka na osnove jemocional'nyh priznakov rechi: dissertatsiya kandidata tekhnicheskikh nauk [Study psychophysiological state based emotional speech features, Ph.D. Thesis]. Khromatidi, A.F. Taganrog.

Mikhaylov, V.T., & Zlatoustova, L.V. (1987) Izmerenija parametrov rechi [Measurements of speech parameters]. Mikhajlov, V.T., Zlatoustova, L.V. Moscow, Radio i svjaz' [Radio and Communication], 167,[1].

Morist, M.U. (2010) Emotional speech synthesis for a radio dj: corpus design and expression modeling: master thesis MTG-UPF dissertation, Barcelona.

Nikishkjan, Je.A. (1986) Tipologija intonacii emotsional'noj rechi [Typology of emotional speech intonation]. Nikishkjan, Je.A. Kiev, Odessa.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. (1957) The measurement of meaning. Urbana. University of Illinois Press.

Perervenko, Ju.S. (2009) Issledovanie invariantov nelinejnoj dinamiki rechi i principy postroenija sistemy audioanaliza psihofiziologicheskogo sostojanija: dissertatsiya kandidata tekhnicheskikh nauk [Investigation of the nonlinear speech dynamics invariants and audio analysis principles of the psychophysiological state, Ph.D. Thesis]. Perervenko, Ju.S. Taganrog.

Rozaliev, V.L. (2009) Modelirovanie jemocional'nyh reakcij pol'zovatelja pri rechevom vzaimodejstvii s avtomatizirovannoj sistemoj: dissertatsiya kandidata tekhnicheskikh nauk [Modeling emotional reactions user when verbal interaction with an automated system, Ph.D. Thesis]. Rozaliev, V.L. Volgograd, VGTU.

Romanenko, R.Ju. (2010) Vejvlet-analiz rechevyh signalov. Uspehi sovremennoj radioelektroniki [Wavelet analysis of speech signals. Advances of modern electronics] R.V. Romanenko. Zarubezhnaja radioelektronika [Foreign electronics]. 12, pp. 51-54.

Schlosberg, H.S. (1941) A scale for the judgement of facial expressions. Experimental Psychology. P. 497-510.

Sidorov, K.V. (2012) Analiz priznakov jemocional'no okrashennoj rechi [Analysis of signs of emotional speech]. Sidorov, K.V., & Filatov, N.N. Vestnik Tverskogo gosudarstvennogo tekhnicheskogo universiteta [Bulletin of the Tver State Technical University]. Issue 20, pp. 26-31.

Sidorov, K.V. (2011) K voprosu ocenki emotsional'nosti estestvennoj i sintezirovannoj rechi po obektivnym priznakam [On evaluation of the natural

- emotion and synthesized speech for objective signs] Sidorov, K.V., Kaljuzhnyy, M.V. Vestnik Tverskogo gosudarstvennogo tekhnicheskogo universiteta [Bulletin of the Tver State Technical University]. Issue 18, pp. 81-85.
- Siging, W. (2009) Recognition of human emotion in speech using modulation spectral features and support vector machines: master of science thesis, Kingston.
- Simonov, P.V. (2011) Leksii o rabote golovnogo mozga: potrebnostno-informacionnaja teorija vysshej nervnoj dejatel'nosti [Lectures on the work of the human brain: the need- information theory of higher nervous activity] Simonov, P.V. Moscow, Nauka, p. 95.
- Simonov, P.V. (1981) Emotsional'nyy mozg [Emotional Brain]. P.V. Simonov. Moscow, Nauka.
- Shpikulyak, I. (2012) Emu ne vse ravno: smartfony smogut razlichat' jemocii [He did not care: smart phones will be able to distinguish between emotions]. Noosfera, IT i elektronika [Noosphere, IT and electronics] - [Electronic resource]. - Mode of access : <http://noos.com.ua/post/3104/> - Access Date December 6, 2012 .
- Sokolov, E.N. (2001) Sfericheskaja model' intellektual'nyh operacij [Spherical model of intellectual operations]. Sokolov, E.N. Psihologicheskij zhurnal [Psychological Journal]. Vol. 22, № 3, pp. 49-56.
- Sokolov, E.N., & Vajtnjavichjus, G.G. (1989) Nejrointellekt: ot nejrona k nejrokompjuteru [Neurointelligence: from neuron to neurocomputer]. Sokolov, E.N., & Vajtnjavichjus, G.G. Moscow, Nauka, p. 238.
- Solovieva, E.S. (2008) Metody i algoritmy obrabotki, analiza rechevogo signala dlja reshenija zadach golosovoj biometrii: dissertatsiya kandidata tekhnicheskikh nauk [Methods and algorithms, speech signal analysis for solving problems of voice biometrics, Ph.D. Thesis]. Solovieva, E.S. Moscow.
- Stanislavskij, K.S. (1959) Moja zhizn' v iskusstve [My Life in Art]. K.S. Stanislavskij. Moscow.
- Starchenko, I.B. (2010) Metody nelinejnoj dinamiki dlja biomeditsinskih prilozhenij [Nonlinear dynamics methods for biomedical applications]. Starchenko, I.B., Perervenko, Ju.S., Borisova, O.S., Momot, T.V., & Izvestija, JuFU. Tekhnicheskie nauki [Proceedings of the SFU. Engineering Special Issue]. 9 (110), pp. 42-51.
- Vartanov, A.V. (2011) Mehanizmy semantiki: chelovek – neuron – model' [Mechanisms of semantics: people – neuron – model]. Vartanov, A.V. Nejrokompjutery: razrabotka, primenenie [Neurocomputers: development and application]. 12, pp. 54-64.
- Vartanov, A.V. (2003) Chto takoe jemocii? 4-h mernaja sfericheskaja model' aspektov perezhivaniya, vyrazheniya, vosprijatija i oboznacheniya emotsij [What is emotion? 4 -dimensional spherical model aspects of the experience, expression, emotion perception and designations] Vartanov, A.V., & Vartanova, I.I. Kul'turno istoricheskij podhod i problema tvorchestva: Materialy vtorykh chtenij pamjati L.S. Vygotskogo [Cultural historical approach and the problem of creativity: Proceedings of the second reading of the memory]: collected papers ed. E.E. Kravtsova, V.F. Spiridonov, & Ju.E. Kravchenko. – Moscow, RGGU, L.S. Vygotskij Fund, pp. 13-29.
- Vartanov, A.V. (2005) Emotsii, motivatsija, potrebnost' v filogeneze psihiki i mozga [Emotions, motivation, needs in the phylogeny of mind and brain]. Vartanov, A.V., & Vartanova, I.I. Vestnik Moskovskogo Universiteta [Bulletin of Moscow University]. Series 14, Psychology, № 3, pp. 20-35.
- Vartanov, A.V. (2001) Chetyrehmernaja sfericheskaja model' emotsij i distantsionnyj rechevoj kontrol' sostojanija cheloveka [Four-dimensional spherical model of emotions and voice remote control of the human condition]. Vartanov, A.V., & Videneeva, N.M. Tezisy dokladov rabochej grupy «Vlijanie informacionnyh tehnologij na nacional'nuju bezopasnost'». 4-ja Ezhegodnaja Konferencija Konsorciuma PrM «Postroenie strategicheskogo soobshhestva cherez obrazovanie i nauku» [Abstracts of the work-group «Impact of Information Technology on the national security», the 4th Annual Conference of the PfP Consortium «Building strategic community through education and research»]. Moscow, p. 35.
- Videneeva, N.M. (2000) Jemocional'nye harakteristiki zvuchashhego slova [Emotional characteristics of the sounding word]. Videneeva, N.M., Khludova, O.O., & Vartanov, A.V. Zhurnal vysshej nervnoj dejatel'nosti [Journal of Higher Nervous Activity]. Vol. 50, Issue 1, pp. 29-43.
- Vilyunas, V.K. (1984) Osnovnye problemy psihologicheskoi teorii jemocij [The main problems of the psychological theory of emotion]. V.K. Viljunas. Psihologija emotsij: teksty [Psychology of emotions] ed. Vilyunas, V.K., & Gippenrejtter, Ju.B. Moscow, Izdatelstvo MGU, pp. 3-26.
- Wundt, V. (1984) Psihologija dushevnyh volnenij [Psychology of excitement]. Wundt, V. Psihologija jemocij: teksty [Psychology of emotions] ed. Viljunas, V.K., & Gippenrejtter, Ju.B. Moscow, Izdatelstvo MGU, pp. 48-63.
- Zlatoustova, L.V. (1957) Tipy jemfaticeskogo udarenija v russkom literaturnom jazyke [Types of emphatic stress in the Russian literary language]. Zlatoustova, L.V. Obshheuniversitetskij sbornik [University collection]. Vol. 117, pp. 107-111.

Работа поддержана РФФИ
проект № 11-06-12036-офи-м-2011.