

СУБЪЕКТИВНОЕ ВОСПРИЯТИЕ И ФОРМАЛЬНЫЙ КОМПЬЮТЕРНЫЙ АНАЛИЗ ЗВУКОВ РУССКОГО ЯЗЫКА¹

© 2016 г. А. В. Варганов*, В. В. Швырев**

* Кандидат психологических наук, старший научный сотрудник
факультета психологии МГУ имени М.В. Ломоносова; Москва;
e-mail: a_v_vartanov@mail.ru

** Студент факультета психологии МГУ имени М.В. Ломоносова; Москва.
e-mail: painspirit@mail.ru

Методом метрического многомерного шкалирования построено субъективное пространство восприятия звуков русского языка (46 фонем, включая варианты гласных и согласных, 17 испытуемых), размерность которого равна 15 (складывается из 4 условно первичных и 11 условно вторичных параметров). Формальный (компьютерный) анализ тех же звуковых образцов с помощью ранее предложенного алгоритма показал, что для их описания достаточно только 4 факторов, которые соответствуют первым 4 субъективным различительным признакам (при усреднении сигнала за все время звучания). Остальные 11 субъективных различительных признаков соответствуют субфонемным квантам – кинакемам (выделяемых по В.Я. Плоткину). Выявленная форма спектрального представления факторов исследуемого набора звуковых сигналов-фонем оказалась сходной с аналогичными характеристиками других наборов звуковых сигналов, которые ранее использовались при исследовании проявления эмоций в звучащей речи и интерпретировались как определенные эмоциональные параметры в четырехмерной сферической модели эмоций.

Ключевые слова: речь, речевые звуки, субъективное восприятие, многомерное шкалирование, эмоции.

Проблема восприятия человеком речевых звуков имеет продолжительную историю и свои традиции. В системе человеческого языка обычно выделяются три подсистемы: фонологическая, лексическая и грамматическая; однако хорошо известно, что за этим троичным делением скрывается иерархически организованная знаковая дихотомия: одноплановая подсистема средств выражения и двуплановая совокупность содержательных, собственно языковых средств, которая дальше распадается на тесно сплетенные, нечетко разграниченные подсистемы лексики и грамматики [14]. Звуковые оболочки различаются по трем уровням содержимого – оболочки морфем, слов и фраз. Поскольку морфема как минимальный знак выступает в речи лишь в составе вышестоящей единицы – слова, ее оболочку можно назвать звуковой лишь с той оговоркой, что “звуковая” не значит “звучащая” [14]: морфема – это только “префабрикат” из фонологических единиц, озвучиваемый при вхождении в звуковую обо-

лочку минимального автономного знака – слова. Звуковая оболочка слова не равна совокупности входящих в нее морфемных оболочек, так как она приобретает дополнительную характеристику – акцентную модуляцию (ударение).

Все существующие теории восприятия речи обычно классифицируются на моторные (наиболее известные представители – Л.А. Чистович, В.Я. Плоткин, В.Н. Сорокин) и сенсорные (Г.М. Фант, Р. Якобсон). Моторная теория исходит из того, что, воспринимая речевой сигнал, слушатель подбирает моторные команды, которые необходимо применить для создания аналогичного сигнала, и использует далее уже не акустическое изображение сигнала, а последовательность этих моторных команд [23]. В сенсорных теориях [12, 18] моторное описание рассматривается как побочное явление. Основной механизм – это сопоставление сигнала с эталоном по акустическим признакам. В последнее же время признается, что как сенсорная, так и моторная теории в крайних своих вариантах не подтверждаются. Современные данные функциональных магнитно-резонансных исследований восприятия речи в

¹ Исследование выполнено при финансовой поддержке РФФИ № 14-06-00327-а и № 11-06-12036-офи-м.

условиях шума [27, 31] выявили активность моторной зоны коры головного мозга, тогда как при хороших акустических условиях активировалась только область слуховой коры [21]. В свою очередь, проводимые в 70-е годы прошлого столетия интенсивные исследования характеристик артикуляторных движений и связанной с этими движениями биоэлектрической активности мышц-артикуляторов доказали, что предположение о стандартности моторных команд не подтверждается [30]. Как и в случае других сложных движений, обнаруживается их высокая вариативность и зависимость от многих конкретных условий исполнения. Например, человек достаточно легко говорит лежа, с сигаретой в зубах и т. д. Конкретные характеристики движений при этом оказываются различными. Таким образом, возникает необходимость интеграции сенсорного и моторного подходов, разработка общей модели механизмов речевосприятия.

Но для экспериментальной верификации разрабатываемых моделей восприятия и порождения речи необходимо выявить параметры звукового сигнала, существенные для распознавания. Основные сведения, используемые для этого, обычно берутся из акустики речи и лингвистики, и касаются свойств речевого аппарата человека и организации языка. С начала прошлого столетия начало формироваться и направление, разрабатывающее автоматические (компьютерные) системы распознавания речи. Долгое время, да и до сих пор, в этих работах общепринятой остается идея, что распознавание связной речи должно обязательно основываться на фонемном или близком к фонемному распознавании. В ее основе лежит допущение о том, что речевой сигнал представлен в сознании слушателя последовательностью дискретных фонетических элементов, а в самом речевом сигнале фонетические элементы присутствуют в неявном виде, для их получения необходима определенная обработка сигнала. При таком фонемном распознавании изолированных звуков речи, звуки рассматривались как сигналы, а фонемы – как названия классов. Решающие функции, найденные, например, в процессе обучения нейросетевых алгоритмов, соответствовали границам между фонемными областями в пространстве выбранных исследователями признаков сигнала. Как выяснилось, хотя автоматическое фонемное распознавание изолированно произнесенных звуков речи (преимущественно исследовались гласные) могло не уступать по надежности фонемному распознаванию их человеком, такая система оказывалась практически неработоспособной в условиях, когда ей предъявлялась для

распознавания связная речь. Поэтому принятая в большинстве работ по автоматическому распознаванию речи точка зрения, что элементами речевого сигнала являются фонемы, которые распознаются последовательно одна за другой, подвергается пересмотру. Исследователи восприятия речи все больше начали склоняться к тому, что фонемы могут распознаваться параллельно и что в качестве фонетического итога выступает слог [1, 26]. При этом сама фонема может формироваться из субфонемных квантов – кинакем [3, 11, 14]. Кинакема неразрывно связана со звуком, служит его реализации и опознанию, но это не звук, а квант деятельности, работы по производству и восприятию звука. Этот квант напрямую связан и с артикуляторными паттернами, но это скорее “центральная моторная команда”, а не мышечный паттерн, который реализует данную команду несколько по-разному в зависимости от условий выполнения движений в соответствии с общими принципами регуляции движений.

В любом случае, необходимо задать некие дифференцировочные признаки, хотя в разных исследованиях выделяется их различное количество и, соответственно, они характеризуются разным содержанием. Так, известно [22, 24, 25, 29], что речевой аппарат имеет достаточно много степеней свободы и, соответственно, может производить несколько десятков различных артикуляторных движений. Но в конкретных языках их существенно меньше, в частности, для русского языка выделяется 11 параметров (см., например, неомоторную теорию В.Я. Плоткина [14]). Идея о наличии у носителей языка системы признаков, по которым воспринимаются и запоминаются фонемы, подтверждается, в частности, существованием у человека представления о субъективном сходстве фонем. Перцептивные признаки фонем можно получить апостериорным путем, в результате анализа реакций испытуемых и использования специальных вычислительных процедур, позволяющих находить минимальную достаточную размерность пространственной конфигурации речевых стимулов на основе информации о субъективном сходстве или различии этих стимулов, полученной экспериментально – методами многомерного шкалирования [32, 28]. Предпринятые ранее исследования восприятия речевых звуков [5, 11] показали, что гласные русского языка субъективно описываются 4-мя параметрами (кинакемными оппозициями: широкорастворность, огубленность, узкорастворность, и осью дифтонгов). Согласные требуют уже 8 признаков, которые также можно было бы охарактеризовать через артикуляторные действия или кинакемы.

Однако общей системы перцептивных признаков для всей звуковой системы конкретного языка до настоящего времени экспериментально не построено. Это связано с большим объемом матрицы различий, необходимой для надежной оценки размерности субъективного пространства: для верификации каждой из осей необходимо как минимум 4–5 объектов сравнения, что задает для 11-мерного пространства, например, необходимость получить матрицу различий для 44–55 объектов сравнения. Это является чрезвычайно трудоемкой задачей не только для экспериментатора, но и для испытуемого. Сделать же вывод о размерности по отдельным небольшим наборам объектов невозможно, т.к. невозможно опровергнуть альтернативную гипотезу, что имеется небольшое число признаков, общих для всех наборов, т.е. возможно не только дополнение этих множеств (с увеличением размерности различительных признаков), но и частичное или полное их пересечение.

С другой стороны, результаты физического анализа речевых звуков, как отмечается в литературном обзоре [15], на современном этапе можно свести к четырем группам объективных признаков и соответствующих им методов, позволяющих различать речевые образцы: спектрально-временные, кепстральные, амплитудно-частотные и признаки на основе нелинейной динамики. Одновременно показано, что на основе одних только простых спектральных характеристик звукового сигнала невозможно правильно распознавать и идентифицировать не только отдельные звуки, но и простые эмоции [16]. Однако и другими методами эффективных решений для распознавания слитной речи или всего спектра эмоциональных выражений в речи до сих пор не найдено. Не спасает дело (несмотря на полученные патенты) и использование принципов бифазной модуляции, широко применяемой в мобильной связи [12, 13]. Тем не менее, недавно проведенный анализ эмоциональной составляющей звучащей речи [3] позволил предложить новый эффективный принцип относительного кросс-частотного амплитудно-вариабельного кодирования информации в звуке. Специфика развиваемого подхода заключается в том, что анализатор речи должен основываться не на статических инвариантах самого звукового сигнала (“звуковых консервах”), с которыми сравнивается текущий сигнал (обычно спектрально представленный), а на характерных изменениях сигнала, отражающих соответствующий паттерн артикуляторных действий. Речь была представлена как многоканальный (разнесенный по частотам) сигнал, в каждой полосе которого возможны независимые быстрые небольшие изменения

амплитуды. Этот метод оказался эффективным для выделения эмоциональной составляющей в речи [4]. В результате показано хорошее согласие выделенных параметров речевого сигнала и субъективного восприятия, описываемого в системе формализованных параметров четырехмерной психофизиологической модели эмоций [6, 9]. Эта модель первоначально была построена с помощью многомерного шкалирования субъективных различий между эмоциональными состояниями, задаваемыми образцами звучащей речи – коротких слов “да” и “нет”, произнесенных с разными эмоциональными выражениями. В дальнейшем она верифицировалась в различных экспериментах (в том числе, при зрительном восприятии выражения лица в статике и динамике, с включенным или исключенным звуковым каналом, а также в семантике русского языка (слов, обозначающих эмоции), и получила хорошее психофизиологическое обоснование. Современные представления о мозге, как о многоуровневой системе управления и отражения, в которой интеллект и аффект представляют две, в некотором смысле автономные, но по необходимости содружественно действующие системы, позволяют понять роль и место эмоций в структуре деятельности. Стойки зрения кодирования информации, наличие разных, но взаимосогласованных уровней требует наличия единых принципов и формы представления входных данных, которые будут описываться в одной и той же системе признаков или параметров регуляции. А это значит, что широкая феноменология аффективных явлений и соответствующих им мозговых механизмов может быть описана в рамках единой модели, которая должна объединить и закономерности отражения эмоций в сознании, и их категоризацию в языке, и их поведенческое выражение, и феномены понимания чужих аффектов. Эти параметры не должны зависеть и от канала выражения эмоций – зрительного или слухового. При этом, в соответствии с теорией кодирования информации в нервной системе номером канала, предложенной Е.Н. Соколовым и формализуемой в виде концептуальной рефлекторной дуги [19, 20], следует допустить, что эта разнородная феноменология может быть объединена с помощью сферической модели следующим образом. Каждое эмоциональное состояние будет описываться в модели точкой в многомерном пространстве, которая должна лежать на поверхности некоторой гиперсферы. Евклидовы оси этого пространства будут представлять нейронные каналы (мозговые механизмы) анализа кодируемого качества состояния воспринимаемых или выражаемых аффектов, а субъективное отражение и

различение в сознании или с помощью языковых (категориальных) средств этих же состояний будет определяться угловой мерой.

В этой разработанной пространственной сферической четырехмерной модели эмоций [6, 9] система первых двух евклидовых осей связана с оценкой ситуации (описывает оценочную функцию эмоций): ось 1 – по знаку (хорошо, полезно или плохо, вредно), ось 2 – по степени информационной определенности (ясность–удивление). В совокупности эта пара осей определяет специфическое качество или модальность эмоции – то, как человек, по мнению слушателя, в том или ином звуке выражает свою оценку ситуации, степень ее информационной определенности и приемлемости для себя. На данной плоскости по кругу эмоциональные стимулы упорядочены в соответствии с эмоциональным тоном по классификации В. Вундта. Третья и четвертая оси совместно описывают побудительную сторону эмоций, то, как человек, по мнению слушателя, собирается действовать. Ось 3 – достижение (монополярная ось), различает пассивные и активные положительные состояния, от спокойного удовольствия (“уверенность”, “радость”) до активного стремления (“любовь”). Ось 4 – оборонительная реакция, активное (агрессия, положительное направление) или пассивное избегание (затаивание, отрицательное направление). По этой оси дифференцируются пассивные и активные состояния, связанные с оборонительной реакцией. Эта пара третьей и четвертой осей также определяет модальность эмоции по В. Вундту, но уже по другому психологическому качеству степени эмоциональной активности (возбуждение – успокоение – угнетение). Угол на данной плоскости определяет последовательность стимулов: ненависть (ярость, гнев) – любовь (стремление) – тревога (страх, ужас). Третий угол четырехмерной гиперсферы, образующейся между первой и второй парами осей координат, будет соответствовать такому психологическому качеству (по В. Вундту), как степень напряжения или разрешения ситуации. Стимулы упорядочены от полной ясности, когда превалирует эмоциональная оценка ситуации, но нет необходимости в действии, до эмоционального напряжения, связанного с неопределенностью ситуации и большой потребностью в действиях.

Такая система параметров хорошо согласуется с мозговыми механизмами эмоций [17, 18]. Так, ось 3 и положительное направление оси 1 отражает работу разных групп нейронов гипоталамуса – побудительных и подкрепляющих, которые хотя и определяют вроде бы сходные положительные эмоциональные состояния, но находятся между собой

в конкурентных отношениях. 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 системой “миндалины – вентральная часть 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 или бегства, соответственно.

Таким образом, формально выделенные в речи параметры [3] хорошо описывают набор эмоционально различающихся речевых образцов, но возникает вопрос, в какой степени эти качества представлены в членораздельной не эмоциональной речи человека? С одной стороны, речь человека на любом языке всегда содержит, кроме информационной, еще и некоторую эмоциональную составляющую. Но, с другой стороны, если уравнивать в звуковых образцах эмоциональное выражение, то какие физические параметры будут их характеризовать? Чем они будут отличаться от эмоциональных параметров? Будет ли человек учитывать эмоциональную составляющую при субъективном сравнении пар звуковых образцов, лишенных явной эмоциональной окраски? Будет ли работать предложенный для эмоциональной составляющей [4] тот же принцип относительного кросс-частотного амплитудно-вариабельного кодирования и для элементов членораздельной речи, в том числе, возможно, субфонемных квантов – кинакем?

Главная цель данной работы – построить субъективное пространство для полного (по возможности) набора всех типичных звуков русского языка, включая как гласные, так и согласные и дифтонги, определив при этом размерность пространства (т.е. число необходимых различительных признаков), и получить интерпретацию осей пространства, что позволит далее сопоставить субъективные и объективные (полученные с помощью компьютерного анализа звука) параметры.

ОСНОВНАЯ ПАРАДИГМА И ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Построение субъективного пространства (метрической модели субъективного восприятия) звуков (фонем) русского языка, включая оцен-

ку размерности (числа необходимых субъективных различительных признаков) и “измерение” каждого из образцов в системе координат выделенных признаков.

2. Измерение физических параметров тех же звуковых образцов – фоном русского языка, которые используются при построении субъективного пространства, на основе алгоритма относительного кросс-частотного амплитудно-вариабельного кодирования, примененного ранее [3] к анализу эмоциональной составляющей речевого сигнала.
3. Сопоставление выделяемых параметров: физически измеренных на данном наборе звуковых образцов и субъективных различительных признаков построенной метрической модели. Такое сопоставление возможно осуществить формально на основании корреляции в наборе используемых звуковых образцов по их координатам – измеренным значениям соответствующих субъективных и объективных параметров.
4. Осуществление интерпретации выделенных субъективных признаков, основываясь на гипотезе о том, что эти субъективные признаки должны соответствовать определенным кинамам русского языка (по В.Я. Плоткину), т.е. описывать данные звуковые образцы так же, как это следует из теории, представленной В.Я. Плоткиным для русского языка [14].
5. Сравнение физического описания (спектрально представленного после применения алгоритма относительного кросс-частотного амплитудно-вариабельного кодирования) исследуемого набора звуковых образцов – фоном русского языка и аналогичного спектрального представления, полученного после применения того же алгоритма к другим наборам звуковых образцов, в частности, к набору коротких слов “Да” и “Нет”, произнесенных с разными эмоциональными выражениями, которые ранее были субъективно специфицированы в четырехмерной сферической модели эмоций [3, 6, 8, 9].

МЕТОДИКА

Стимульный материал

В аудиоредакторе *GoldWave* v5.12 установленном под платформой *Windows 7* на *PC HP pro 3500 mt* с помощью гарнитуры *Logitech USB Headset H530* мужчиной в возрасте 30 лет были записаны (оцифрованы с частотой 44100 Гц, 16-bit, Mono) 46 звуков русского языка, включая 6 гласных звуков: [а] [и] [о] [у] [ы] [э], 36 согласных звуков: [б]

[бь] [в] [вь] [г] [гь] [д] [дь] [ж] [й] [з] [зь] [к] [кь] [л] [ль] [м] [мь] [н] [нь] [п] [пь] [р] [рь] [с] [сь] [т] [ть] [ф] [фь] [х] [хь] [ц] [чь] [ш] [щ] и 4 дифтонга [йэ], [йо], [йу], [йа].

Далее данные звуки были нарезаны в аудиоредакторе *GoldWave* v5.12 на отдельные файлы формата *.wav, длительность которых составляла от 135 до 500 мс.

Полученные файлы загружались в программный комплекс “Рабочее место психолога Практика – МГУ” (РМП Практика) версии: 2014 (001.000.001.034) совместимый с *Windows 7* в директорию “*ScalMake* – Конструктор задач на шкалирование”.

Испытуемые

Испытуемые 17 человек (трое мужчин и четырнадцать женщин), студенты факультета психологии МГУ имени М.В. Ломоносова в возрасте 18–34 лет.

Процедура эксперимента

Эксперимент состоял из двух серий: короткого тестового ознакомительного задания (28 предъявлений), в котором были представлены звуки [о], [э], [б], [п], [с], [м], [щ], [т], предъявляемые в фиксированном порядке, и основного задания, в котором были представлены все 46 звуков, предъявляемых попарно в случайном порядке (всего 1035 пар, предъявляемых по одному разу).

Стимулы в паре предъявлялись последовательно, один за другим, с межстимульным интервалом 300 мс с программными параметрами предъявления каждого стимула 700 мс. Фиксировалась субъективная оценка величины различия двух стимулов по 9-балльной шкале от 1 до 9, шкала выводилась на экран монитора. Следующая пара стимулов предъявлялась только после ответа на предыдущую пару. Стимулы предъявлялись через наушники.

Инструкция испытуемому (выводилась на экран монитора).

В данном задании Вам будут предъявляться фонемы русского языка. Звуки будут предъявляться попарно – один за другим с паузой в 0.5 секунды. Ваша задача заключается в том, чтобы, прослушав оба звука, оценить субъективное различие между ними по 9-балльной шкале. Чем больше кажущееся различие, тем большим числом Вы его оцениваете. “1” – минимальное различие, “9” – максимальное различие. Для ответа используйте “мышь” или клавиши клавиатуры. Обязательно давайте ответ после предъявления КАЖДОЙ

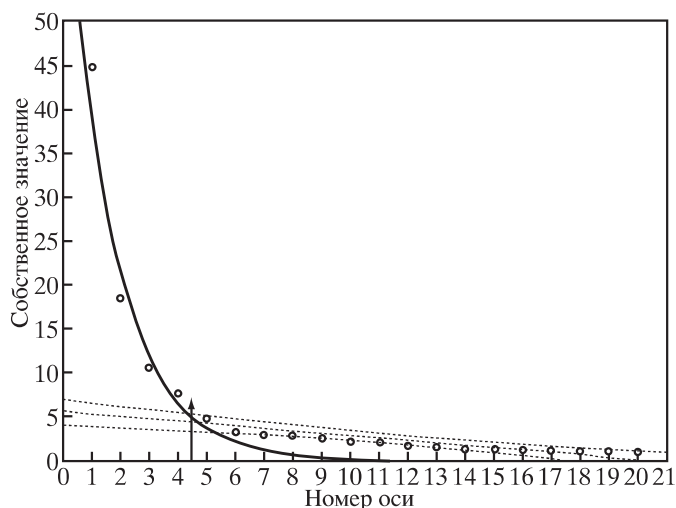


Рис. 1. Оценка размерности получаемого факторного пространства физических описаний набора звуковых образцов на основе критерия “каменистой осыпи”.

пары звуков. Пока Вы не дали ответ, следующая пара звуков предъявляться не будет.

Процедура анализа звукового сигнала.

Спектральный показатель вычислялся с использованием специальной (авторской) программы по следующему алгоритму:

Для звукового фрагмента с помощью стандартных средств – быстрое преобразование Фурье со сглаживанием в минимальном скользящем окне – вычислялась последовательность мгновенных спектров мощности сигнала (в диапазоне от 0 до 7000 Гц с разрешением 50 Гц). На основе последовательности мгновенных спектров в скользящем окне (размером 20 мс и шагом сдвига 10 мс) вычислялся показатель микро-вариативности (стандартное отклонение) амплитуды (квадратного корня от мощности) на каждой частоте, который нормировался по частотному диапазону. В результате для всего периода звучания для каждого образца с шагом 0.01 с были получены 141-мерные нормированные вектора данных.

Далее ко всему полученному массиву данных (для всех образцов по всему времени звучания) применялся факторный анализ с целью оценки размерности (числа необходимых факторов, описывающих большую часть изменения сигнала) и редукции данных – измерения значений факторов, представления каждого звукового образца вектором меньшей размерности (предположительно, четырехмерного).

В случае необходимости, для вычисления интегральной оценки всего звукового образца использовалось простое усреднение предыдущего

показателя по всему интервалу звучания и получения одного усредненного вектора для каждого звукового образца.

Методы анализа данных о субъективных различиях.

Полученные от разных испытуемых матрицы субъективных различий первоначально анализировались с помощью факторного анализа на предмет выявления специфики “точек зрения” испытуемых, а затем разные варианты матриц различия анализировались метрическим методом многомерного шкалирования [32]. Массив пересчитанных по вышеописанному алгоритму значений звукового сигнала подвергался факторному анализу.

РЕЗУЛЬТАТЫ

Результаты факторного анализа формально выделенных спектральных сигналов

Размерность получаемого факторного пространства физических описаний набора звуковых образцов оценивалась по распределению собственных значений, которое представлено на рис. 1. Доля описываемой дисперсии для 4 факторов равна 57.73 % от общей дисперсии (дисперсия имела место не только между исследуемыми образцами, но и внутри каждого образца в связи с изменением звука во времени). На основании этих данных размерность пространства должна быть оценена как равная 4.

Факторные нагрузки для четырех выделенных факторов после вращения по методу Варимакс в их спектральном представлении даны на рис. 2. Для сравнения на этом же рисунке представлены и 4 аналогичных параметра, которые были выделены ранее таким же методом на других наборах звуковых образцов, в частности, на наборе коротких слов “Да” и “Нет”, произнесенных с разными эмоциональными выражениями. Видно хорошее соответствие между выделенными параметрами, описывающими фонемы русского языка, и параметрами, описывающими эмоции в ранее обоснованной четырехмерной сферической модели эмоций [4, 7, 9]:

- выделяемый фактор 1 хорошо коррелирует с фактором 1 эмоционального пространства ($r = 0.8599$), который ранее интерпретирован как ось знака эмоций (плохо–хорошо);
- выделяемый фактор 2 хорошо коррелирует с фактором 2 эмоционального пространства ($r = 0.7478$), который ранее интерпретирован

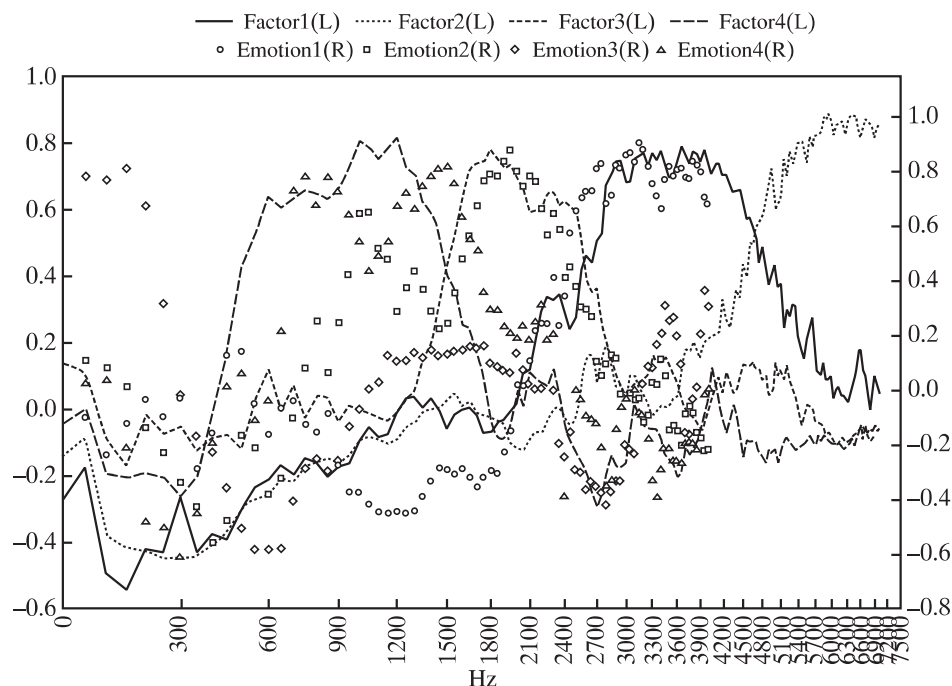


Рис. 2. Спектральное представление полученных факторных нагрузок, которые описывают физические свойства звука, характерного для данного набора образцов фонем русского языка (значения по левой оси) показаны разными линиями. Различными маркерами отмечены полученные ранее аналогичные показатели, выявленные на другом наборе – слов “Да” и “Нет”, произнесенных с разными эмоциональными выражениями и специфицированных в четырехмерной сферической модели эмоций (в связи с разным масштабом, эти значения даны по оси справа). Горизонтальная ось частоты дана в логарифмическом масштабе.

как ось информационной определенности (удивления);

- выделяемый фактор 4 хорошо коррелирует с фактором 4 эмоционального пространства ($r = 0.8633$), который ранее интерпретирован как ось отвержения или обороны (пассивного отвержения – страх, затаивание, или активного отвержения – агрессия, ярость).

Только в одном случае – для оставшегося эмоционального параметра 3 (интерпретированного ранее как ось притяжения) совпадение в низкочастотной области плохое ($r = 0.0060$). Это объясняется тем, что ранее субъективно специфицированные эмоциональные образцы были записаны с частотой, не превышающей 8 кГц, поэтому был возможен анализ их спектра только до 4 кГц, тогда как в данном исследовании анализировался более широкий диапазон. Именно этот параметр имеет существенное продолжение в высокочастотной области, которая не учитывалась ранее при выделении соответствующего эмоционального параметра. Таким образом, и данную ось все же можно интерпретировать как эмоциональный параметр 3 (притяжение).

В результате были получены также и значения выделенных факторов, которые будут проанали-

зированы совместно с данными субъективного оценивания тех же звуковых образцов.

Результаты эксперимента по многомерному шкалированию звуков русского языка.

Первоначально для оценки “точек зрения” испытуемых результаты шкалирования объектов (звуков русской речи или фонем) каждого испытуемого были сопоставлены между собой с помощью факторного анализа. Обнаружено, что эти оценки, хотя и содержат большой шум, но достаточно хорошо согласованы: генеральный фактор описывает 46.1% общей дисперсии, второй по значимости фактор составляет только 7.3%, а остальные – менее 5% дисперсии. Это свидетельствует, что несмотря на имевшиеся различия (по полу, возрасту и проч.), у всех испытуемых субъективные впечатления при сравнении данных звуковых стимулов в значительной мере совпали.

В дальнейшем анализировались следующие варианты матриц различия:

- 1) Усредненная матрица различий по всем 17-ти испытуемым.
- 2) Усредненные матрицы различий, построенные по отдельности для двух групп наибо-

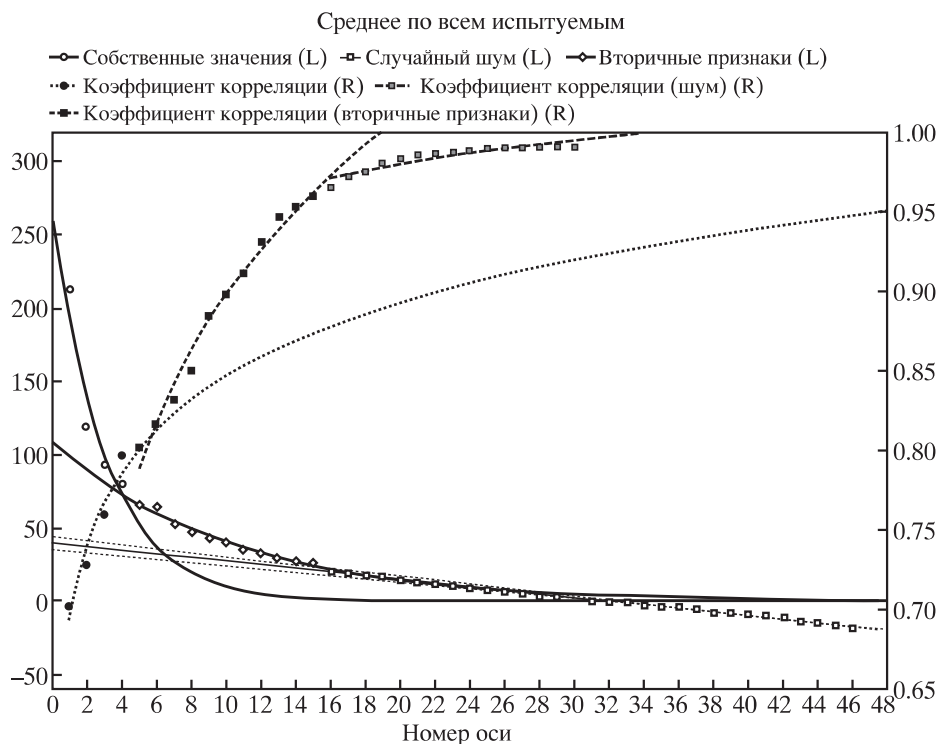


Рис. 3. Распределение собственных значений и коэффициентов корреляции евклидовых расстояний в моделях разной размерности и субъективных оценок.

лее “типичных” испытуемых, отобранных на основе факторных нагрузок после варимакс-вращения для выделения этих “точек зрения”.

- 3) Матрица, построенная по значениям факторов (после приведения их к масштабу с минимумом = 0 и максимумом = 9) – матрица для генерального фактора.
- 4) Матрицы для значений двух факторов, выделяющих “точки зрения” двух групп испытуемых.

В результате оказалось, что размерность получаемого пространства по распределению ее собственных значений и графику стресса (по коэффициентам корреляции евклидовых расстояний в модели соответствующей размерности и исходных субъективных различий) для всех исследованных вариантов оказалась сходной. На рис. 3 для примера представлены результаты оценки размерности субъективного пространства для матрицы различий, усредненной по всем испытуемым (первый вариант). На графиках выделены три периода – кружки отмечают первые 4 размерности, ромбики и треугольники – размерность с 5 по 16 (11 отсчетов), а квадратики – оставшиеся размерности. Аппроксимация линейной функцией выполнена на основе подгонки только отрица-

тельных собственных значений, соответствующих в метрической модели исключительно шуму.

На этом рисунке видно, что распределение коэффициентов корреляции исходных субъективных оценок с евклидовыми расстояниями в моделях разной размерности хорошо аппроксимируется логарифмической функцией, тогда как собственные значения – экспонентой и линейной функцией. Известно, что матрица случайных чисел имеет собственные значения, график которых аппроксимируется прямой линией (при этом примерно половина значений отрицательны, т.к. за счет случайности нарушается требование существования треугольника в евклидовом пространстве), тогда как значимые различительные признаки характеризуются экспоненциальной функцией. Это и является основанием для оценки размерности, отделения значимых признаков от осей, описывающих случайный шум. Однако на вышеприведенном и других аналогичных (для других вариантов матриц) графиках можно обнаружить, что имеются две группы значимых признаков, каждая из которых аппроксимируется своей экспоненциальной функцией, имеется явный перегиб. На некоторых графиках это заметнее, чем на других. Этому соответствует и распределение коэффициентов корреляции: хотя вид этих графиков схож – во всех случаях это

логарифмическая функция, однако параметры для этих трех отрезков размерности различаются, имеют характерный “перегиб”. Таким образом, эти результаты позволяют заключить, что во всех случаях в анализируемых матрицах различий заключены три типа данных: 1) наиболее выраженные 4 признака, которые можно назвать первичными; 2) 11 вторичных содержательных признаков; 3) размерности, описывающие случайные ошибки (шум) в данных. При этом в некоторых случаях разделение 4 первичных и 11 вторичных признаков выражено сильнее, а в других – слабее, но все равно заметно. В итоге общую размерность субъективного пространства речевых звуков русского языка можно считать равной 15.

Поскольку по всем вариантам анализа точек зрения испытуемых была получена сходная картина, в дальнейшем анализировалась только усредненная по всем 17 испытуемым матрица различий (первый вариант). Полученная конфигурация точек-стимулов (звуков речи) в 15-мерном евклидовом пространстве проверялась на сферичность и подвергалась процедуре вращения. При этом даже в пространстве первых 4 осей конфигурация точек не заполняла все пространство непрерывно, а образовывала сферическую поверхность: средний радиус = 3.395546, стандартное отклонение радиуса = 0.376302, коэффициент вариации = 0.110822 (11.08%). Для 15-мерного решения сферичность оказалась еще более выраженной: средний радиус = 4.733096, ст. отклонение радиуса = 0.151731, коэффициент вариации = 0.032058 (3.21%).

Сопоставление объективных и субъективных различительных параметров.

Полученная конфигурация точек субъективного пространства подвергалась процедуре ортогонального вращения с целью проверки уточненной гипотезы с помощью специальной (авторской) программы: подбирались углы поворота осей координат полученного субъективного пространства для достижения наилучшего соответствия с заданными на основании гипотезы моделью (по критерию минимизации квадрата разницы координат). Проверялась гипотеза, что первые 4 оси субъективного пространства могут соответствовать средним арифметическим по времени звучания 4 параметрам (факторам), которые были выделены при формальном анализе звуковых образцов и соответствуют параметрам оценки эмоций в четырехмерной сферической модели. Другие 11 осей субъективного пространства должны

описывать различия исследуемых звуков (фонем) в соответствии с 11 кинакемами (субфонемными квантами), которые выделяются В.Я. Плоткиным [14]: модель в этом случае была составлена по схеме кинакем, ее 11 оппозиций с необходимой нормировкой. Интерпретация основана на расположении всех использованных стимулов в системе выделяемых осей пространства.

В результате вращения удалось получить хорошее соответствие координат субъективного пространства и теоретической модели: итоговая интерпретация осей субъективного пространства различения звуков и соответствующие коэффициенты корреляции представлены в табл. 1.

Из таблицы видно, что все величины коэффициентов корреляции высоки и значимы, что говорит о хорошем соответствии полученной конфигурации с теоретически предсказанной в соответствии с гипотезой.

Таким образом, выделяемые в субъективном восприятии 15 параметров метрической модели получили свою интерпретацию: первые 4 оси могут быть интерпретированы как физические измеренные параметры, которые схожи с параметрами, ранее выделенными для описания эмоциональной составляющей речи человека [4] (1 – ось знака, 2 – ось информационной определенности или удивления, 3 – ось притяжения, 4 – ось отвержения, активного или пассивного); остальные 11 осей могут быть интерпретированы как динамические паттерны изменения звука в связи с порождающими их артикуляторными паттернами – кинакемами (субфонемными квантами). Распределение исследуемых стимулов-звуков по выделенным субъективным осям также подтверждает данную интерпретацию осей. Для примера на рис. 4 показано распределение стимулов по оси №11, соответствующей кинакемной оппозиции “сонорность”.

При этом обнаруживается, что первые 4 параметра субъективного восприятия звуков (фонем) русского языка хорошо совпадают с объективно выделяемыми усредненными параметрами звука, которые, в свою очередь, задают (исходя из интерпретации их спектральных характеристик) общие эмоциональные качества. И это несмотря на то, что при субъективном шкалировании не было задачи оценивать эти звуки по эмоциональной характеристике, и испытуемые не отмечали, что различают их по этим признакам.

Таким образом, оказалось, что анализ объективно полученных параметров, усредненных по всему диапазону звучания звуков (фонем), позволяет

Таблица 1. Корреляции координат объектов в субъективном пространстве и их теоретически заданных координат (все коэффициенты значимы на уровне $p < 0.05$)

Теоретически задано	Содержательная интерпретация	№ оси субъективного пространства	Коэффициент корреляции
Физические параметры звука			
<i>Mean FACTOR1</i>	Физически параметр звука 1, выделенный на основе алгоритма относительного кросс-частотного амплитудно-вариабельного кодирования в данном наборе образцов	1	0.84
<i>Mean FACTOR2</i>	Физически параметр звука 2, выделенный на основе того же алгоритма (см. выше)	2	0.65
<i>Mean FACTOR3</i>	Физически параметр звука 3, выделенный на основе того же алгоритма (см. выше)	3	0.83
<i>Mean FACTOR4</i>	Физически параметр звука 4, выделенный на основе того же алгоритма (см. выше)	4	0.74
Кинакемые оппозиции (изменение звука при выполнении соответствующих артикуляций)			
Огубленность	Изменения звука в связи с изменением горизонтальной конфигурация ротового резонатора (сдвинутого вперед или назад), удлиненного с помощью вытянутых губ. Гласные заднего ряда лабиализованные ([y] [o])	5	0.78
Узкорастворность	Изменения звука в связи с растягиванием рта вширь определяет степень подъема языка к небу: гласные верхнего ([и], [ы], [у]) и среднего подъема ([э], [о])	6	0.56
Широкорастворность	Изменения звука в связи с раскрытием рта (опусканием челюсти) с варьированием узкого и среднего растворов определяет место подъема языка: гласные нижнего ([а]) и среднего подъема ([э], [о]).	7	0.65
Палатализация	Изменение звука при произношении мягких согласных ([бь], [вь], [гь], [дь], [зь], [пь], [фь], [кь] [ть] [сь], [мь], [нь], [рь], [ль], [хь], [чь], [шь], [жь], [й]), когда тело языка сдвигается вперед, а средняя часть спинки языка приподнимается к твердому небу. Палатализация считается дополнительной артикуляцией: она накладывается на основную, связанную с образованием преграды.	8	0.52
Активная предцентральность	Изменения звука в связи с использованием нижней губы как активного артикулятора. Пассивный артикулятор – верхняя губа ([б] [бь] [п] [пь] [м] [мь]) или верхние зубы ([в] [вь] [ф] [фь]).	9	0.75
Активная зацентральность	Изменения звука в связи с использованием задней части спинки языка – активный артикулятор. Пассивный артикулятор – средняя часть твердого неба ([гь] [кь] [хь]); или задняя часть твердого неба ([г] [к] [х]).	10	0.68

Таблица 1 (окончание)

Теоретически задано	Содержательная интерпретация	№ оси субъективного пространства	Коэффициент корреляции
Пассивная зацентральность	Изменения звука в связи с тем, что кончик языка вместе со спинкой приподнимается к верхним зубам и альвеолам ([ж] [ш] [щ] [чь]) или же средняя часть спинки языка сближается с твердым небом ([й]).	11	0.51
Смычность	Изменения звука в связи с тем, что в гортани происходит раскрытие–закрытие голосовой щели, помогающее остановить колебания голосовых связок и увеличить ротовое давление. Кроме того, поднимается язычок (маленький конусообразный придаток небной занавески – увула) и закрывает носовой проход. В итоге за смыканием органов артикуляции следует их резкое размыкание, сопровождаемое сильным толчком воздуха ([б] [бь] [д] [дь] [г] [гь] [п] [пь] [т] [ть] [к] [кь]), смычное начало артикуляции и щелевое завершение([ц] [чь]);	12	0.78
Проточность	Изменения звука в связи с тем, что одни органы артикуляции образуют смычку, а другие образуют проход для воздушной струи ([л] [ль] [м] [мь] [н] [нь]).	13	0.77
Шумность	В шумных звуках [б], [б'], [в], [в'], [г], [г'], [д], [д'], [ж], [ж'], [з], [з'] шум возникает в результате преодоления преграды воздушной струей.	14	0.74
Сонорность	Изменения звука в связи с тем, что в ротовой полости создается преграда, но воздушная струя находит относительно свободный выход наружу и шум не образуется. Воздух устремляется либо через нос ([м], [м'], [н], [н']), либо в проход между боковыми краями языка и щеками ([л], [л']). Отсутствие шума может быть связано с мгновенностью преграды ([р], [р']) или с достаточно широким характером самой щели ([j]).	15	0.68

объяснить только 4 субъективные оси пространства, но для описания всей системы субъективных оценок требуется кроме этих 4, еще 11 дополнительных параметров, различающих данные звуки. Каковы же физические основы для данных субъективных признаков? Как показал анализ, эти дополнительные параметры формируются на основе динамики изменения звука в процессе его звучания. Полученные значения 4 “объективных” факторов для каждой из фонем позволили исследовать, как они изменяются в процессе звучания.

Обнаружено, что эти изменения не случайны, а носят систематический характер и определяются координированным изменением значения по выделенным факторам. Корреляционный и факторный анализ изменения значений этих параметров для каждой из фонем в отдельности показал наличие существенных корреляций, специфичных для каждой из фонем, а также наличие общих принципов изменения фонем, обусловленных наличием 11 типов действий (кинаем). Для этого все исследуемые звуки были распределены на 11

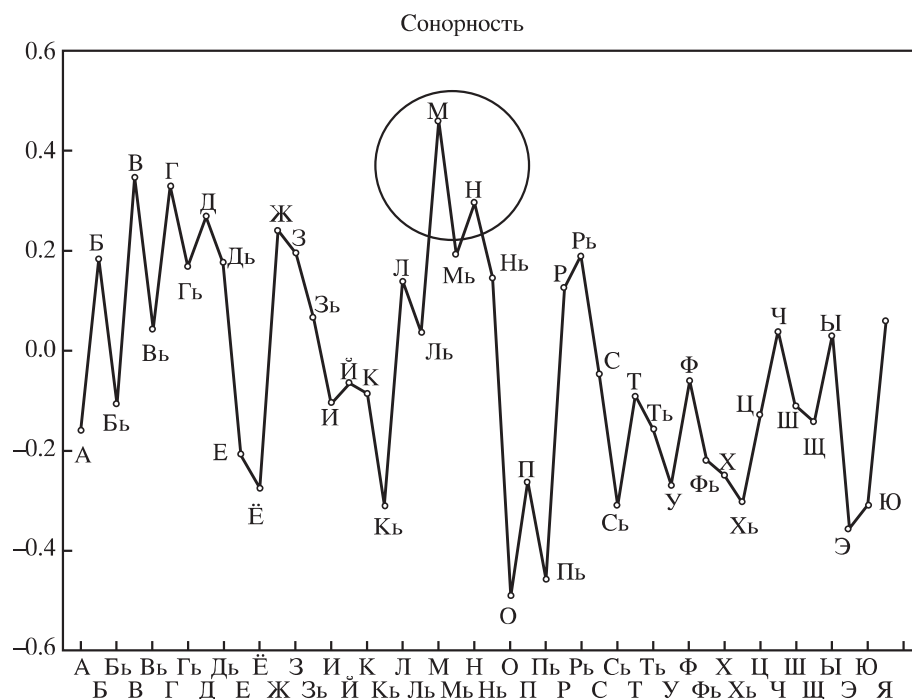


Рис. 4. Распределение исследуемых звуков (фонем) русского языка по выделенному субъективному параметру (ось 11) – кинакемной оппозиции “сонорность”.

групп по доминированию определенной кинакемы (вклада соответствующего ей субъективного признака), которые затем и были обработаны факторным анализом. В результате обнаружено, что размерность во всех случаях не превышает 2, поэтому далее каждая из групп была охарактеризована собственным паттерном изменения в этом двухфакторном пространстве. На рис. 5 для примера представлена траектория изменения звуков “М” и “Н”, в которых доминирует кинакемная оппозиция “сонорность”, в системе вторичных осей, полученных с помощью вторичного факторного анализа и выделения двух вторичных факторов. Оси на рисунке соответствуют этим вторичным факторам, задающим двумерную плоскость в четырехмерном пространстве. Для данной кинакемной оппозиции характерно одновременное изменение звука по параметрам 2 с плюсом (факторная нагрузка 0.81) и 3 с минусом (факторная нагрузка -0.93) (горизонтальная ось) и синхронное изменение по параметрам 1 (факторная нагрузка 0.76) и 4 (факторная нагрузка 0.98) с плюсом (вертикальная ось).

Таким образом, полученные данные позволяют заключить, что выделенные ранее для эмоциональной составляющей речи человека [4] 4 параметра, так же хорошо описывают и членораздельную речь. Причем параметры, усредненные по всему диапазону звучания стимула, отражают, по-видимому, в большей степени эмо-

циональную (просодическую) составляющую, а динамика специфического изменения этих параметров, связанных с типами артикуляторного действия (кинакемами), характеризует и качество звуковых единиц (фонем), что подтверждается данными их субъективного шкалирования.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В недавно проведенном исследовании звучащей речи человека [3] был предложен новый эффективный метод автоматического распознавания эмоций по речевому сигналу, основанный на четырехмерной сферической модели эмоций и принципах кодирования информации в нервной системе [4]. Полученные параметры (факторы) можно охарактеризовать как бимодальные спектральные фильтры. Фактор 1 имеет основной максимум в области 3000 Гц и вспомогательный – на 500 Гц. Он определяет изменение звукового сигнала по оси знака эмоций – чем более вклад данного компонента по сравнению с другими компонентами, тем положительнее (лучше, полезнее) оценивается объект высказывания. Фактор 2 имеет два близких максимума на частотах 1000 и 1750 Гц. Он определяет степень информационной неопределенности ситуации – удивление в противоположность уверенности (спокойствию). Фактор 3 имеет самые широко разнесенные максимумы – в

низкочастотной области (около 150 Гц) и высокочастотной области (3500 Гц при анализе в диапазоне до 4000 Гц). Он характеризует притяжение (любовь), при этом для набора слов “нет” он сопровождается отсутствием активного отвержения, а для набора “да” – положительной оценкой (знаком). Фактор 4 имеет близкие максимумы на 600 и 1500 Гц, по конфигурации близок к фактору 2, но сдвинут относительно него в низкочастотную область, попадая своими максимумами в его локальные минимумы. Этот компонент соответствует характеру отвержения, определяет, будет ли реакция активной (агрессивной) или пассивной (страх, бегство).

Результаты применения этого же алгоритма для анализа фонем в целом подтвердили продуктивность предлагаемого антропоморфного подхода к разработке технических систем, в частности, методам обработки речевого сигнала и представления данных. Однако эти данные были получены только для относительно узкого частотного диапазона – до 4 кГц. В данном же исследовании этот принцип и алгоритм вычисления сигнала применялся для образцов в более широком диапазоне – до 7 кГц. При этом использовался совершенно другой материал, в котором эмоциональная составляющая была по возможности нивелирована. Речевые звуки произносились с одной и той же интонацией. Тем не менее в результате такого объективного анализа были получены факторы, которые достаточно хорошо совпали с ранее выявленными “эмоциональными” параметрами. При этом оказалось, что три фактора согласуются очень хорошо, т.е. получена сходная конфигурация низкочастотной области (до 4 кГц), хотя они продолжают и в высокочастотной области (выше 4 кГц). Только один фактор (номер 3, соответствующий эмоции притяжения) в низкочастотной области оказался не слишком схожей конфигурации, и именно для него обнаружена существенная высокочастотная составляющая, выходящая за пределы 4 кГц. Можно отметить, что проведенное исследование позволяет уточнить ранее выделенные параметры эмоционального сигнала, хотя такая задача и не ставилась. То, что фактор 3, несмотря на выявленные различия, все же имеет отношение к эмоции притяжения, можно обосновать наблюдением, что высота основного тона речи матери, обращаемой к своему ребенку, повышается. По-видимому, именно наличие высокочастотной области “фильтра”, выделяющего эмоциональное значение в звуке, характеризует его качество как “притяжение”.

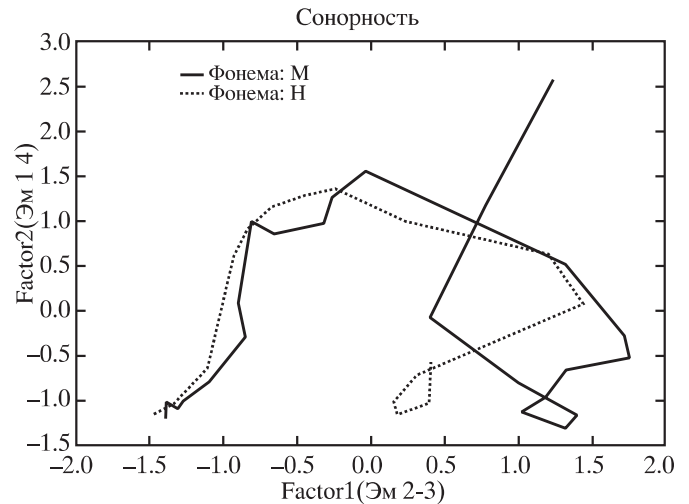


Рис. 5. Динамика (траектория изменения) формальных параметров звукового сигнала для группы звуков (“М” и “Н”) с доминированием кинакемной оппозиции “Сонорность” (справа) в пространстве двух вторичных факторов. Начальные точки отмечены маркерами.

Таким образом, первые четыре выделенных оси субъективного пространства, которые по своим физическим параметрам оказались схожими с параметрами, описывающими изменения эмоциональных звуковых сигналов, позволяют человеку выделять эмоциональное содержание в речи:

Ось 1 пространства фонем соответствует оси 1 модели эмоций – знаку эмоций (хорошо–плохо);

Ось 2 пространства фонем соответствует оси 2 модели эмоций – информационной определенности или удивления (ясность–неопределенность);

Ось 3 пространства фонем соответствует оси 3 модели эмоций – притяжению (любви);

Ось 4 пространства фонем соответствует оси 4 модели эмоций – отвержению (пассивной обороне, страху, затаиванию – активной обороне, агрессии, ярости).

Сходство выделенных физических параметров звука для нейтрально эмоциональных, на первый взгляд, речевых звуков с параметрами эмоциональной составляющей речи человека [4], позволяет объяснить ряд феноменов, описываемых в фоносемантике [10]. В частности, произнесенное слово, состоящее из последовательности определенных звуков, кроме семантического значения, несет в себе и определенную эмоциональную нагрузку. Одно и то же содержание, передаваемое, например, словами-синонимами, может иметь разную эмоциональную окраску в зависимости от своего конкретного звукового состава.

Этот эффект хорошо известен в фоносемантике и применяется на практике (например, при подготовке речи спичрайтерами), однако обоснования его механизмов до настоящего времени не было. Полученные в данном исследовании результаты позволяют их понять и более тонко учитывать, а также подойти к разработке соответствующих программных средств. Конечно, в рамках одной статьи невозможно дать полный сопоставительный анализ полученных результатов, затрагивающих обширную тему фонетического символизма, что потребовало бы подробного сравнения с результатами А.П. Журавлева и метода семантического дифференциала, равно как и решения ряда проблем, связанных с интерпретацией факторов, и с тем, как вообще следует понимать эмоции. В данном подходе также не учитывается ряд важных просодических характеристик звучащей речи – интенсивность звучания и темпоральные характеристики, что представляется задачей дальнейших исследований.

Таким образом, полученные результаты не только хорошо согласуются с известными феноменами и могут способствовать их объяснению, но позволяют поставить вопрос об уточнении ранее выделенных физических параметров, кодирующих эмоциональное содержание речевых высказываний, в частности, за счет расширения частотного диапазона.

Полученные результаты также позволяют обсуждать и роль артикуляторных действий в восприятии речевых звуков, их мозговых механизмов. Факт интерпретации субъективных признаков восприятия в соответствии с выделенными В.Я. Плоткиным кинакемами, свидетельствуют скорее в пользу неомоторной теории речевосприятия. Функция звука как продукта артикуляции и объекта аудиоперцепции – экстернализация речевого сигнала и перенос его от человека к человеку. Можно согласиться с В.Я. Плоткиным [14], что путь сигнала начинается не во рту и завершается не в ухе – и отправителем, и получателем сигнала является мозг. Три квантовые фонологические единицы – кинакема, фонема, слог – в своей совокупности составляют иерархически организованный механизм, функциональное назначение которого – рождение звуковых оболочек для системы знаковых единиц языка. Кинакема есть языковая единица, придающая лингвистическую организацию нервно-мышечной деятельности говорящего и соответствующей мышечно-нервной деятельности слушающего. Для русского языка Р. Якобсон в своих исследованиях выявил 12 оппозитивных пар различительных признаков,

а В.Я. Плоткин [14] – 11. Наше исследование подтверждает правоту выделения именно 11 кинакем, поскольку это доказывается обнаруженной размерностью субъективного пространства.

Полученные в данном исследовании результаты позволяют перейти к обсуждению проблемы возникновения человеческого языка. Приведенный в [2] обзор основных теорий его происхождения показывает, что разные теории “выхватывают” разные стороны и моменты, которые, однако, могут быть соединены в одну общую систему, которую можно предложить, исходя из полученных результатов. Тот факт, что система речевых звуков современного русского языка содержит и базируется на эмоциональных параметрах, обнаруживаемых как в речи человека [3], так и в звуках обезьян [8] доказывает, что язык человека возник как усложнение эмоциональной коммуникации, что соответствует теории междометий (Руссо и его последователи). При этом система эмоциональных голосовых (и не голосовых) сигналов является врожденной – общей как минимум для всех приматов. Так, звуковые эмоциональные сигналы обезьян макак-резусов воспринимаются человеком в той же системе параметров, что и человеческие эмоции, по-видимому, эти обезьяны и испытывают тот же спектр эмоциональных переживаний, что и человек [7]. Все это объясняет, почему ребенку не требуется специальных усилий, чтобы понять эмоциональную основу обращенной к нему речи, но ему приходится учиться произвольно изменять эти сигналы, поскольку динамика изменения эмоционального выражения определяет специфику фонем.

С другой стороны, несмотря на возможное существование единой эмоциональной основы всех языков (данное положение еще требует прямого экспериментального подтверждения, однако общий даже с обезьянами способ кодировки эмоций делает такое предположение весьма вероятным) разные языки существенно различаются по составу фонем и кинакем. Поэтому формироваться в филогенезе и усваиваться в онтогенезе каждый язык может совершенно независимо. Специфические паттерны изменения голосового эмоционального выражения могут формироваться и закрепляться в социуме, что свидетельствует в пользу необходимости объединения коммуникации и мышления в единую речемыслительную систему, на чем базируется гипотеза Н.А. Барулина – “инструментальная” семиотическая система образуется путем закрепления бывших (еще у узконосых обезьян) так называемых “*ad-hoc*-сигналов” и передачи их “по наследству” посредством

обучения и подражания, так же, как умения изготавливать орудия труда [1, 2].

Полученные результаты не позволяют согласиться с тем, что такое объединение возникло в результате мутации, поскольку у всех приматов уже имеется одинаковая эмоциональная основа. Это противоречит гипотезам Д. Бекертона, С. Пинкера и Т. Дикона — для возникновения языка не нужен новый “субстрат”. Но можно согласиться, что кроме объединения мышления с речью, должны были появиться способности, во-первых, к звукоподражанию, а во-вторых, к необыкновенно тонкому регулированию дыхания, без которого невозможна речь (на чем также настаивает Н.А. Барулин [1]). Кроме того, полученные результаты противоречат теориям жестов и мимики как возможной основы происхождения звукового языка, поскольку, как оказалось, современный звуковой язык основывается на звуковой эмоциональной системе, столь же древней, как и система проявления эмоций в мимике и жестах.

ВЫВОДЫ

1. Построенное субъективное пространство восприятия звуков русского языка (метрическая модель) имеет общую размерность, равную 15, которая складывается из 4 условно первичных и 11 условно вторичных параметров (евклидовых осей пространства).
2. Формальный (компьютерный) анализ исследованных звуковых образцов с помощью алгоритма относительного кросс-частотного амплитудно-вариабельного кодирования позволил обнаружить, что все разнообразие изменения звука описывается четырьмя факторами, которые по своим физическим (спектральным) характеристикам оказались сходными с параметрами, описывающими эмоции в звучащей речи (осями четырехмерной сферической модели эмоций).
3. Первые 4 параметра (субъективных признаков или евклидовых осей пространства) различают исследуемые речевые звуки точно так же, как и формально выделенные параметры (факторы) речевого сигнала, которые, в свою очередь, соответствуют осям четырехмерной сферической модели эмоций (оси 1 знака эмоции, оси 2 информационной неопределенности или удивления, оси 3 притяжения и оси 4 отвержения).
4. Дополнительные 11 параметров (субъективных различительных признаков или евклидовых осей пространства) могут быть хорошо интерпретированы, исходя из соответствия с теоретически выделяемыми единицами — 11 каноническими оппозициями для русского языка, которые, в свою очередь, объективно определяются спецификой траектории изменения сигнала в четырехмерном пространстве выделенных параметров (интерпретируемых как параметры эмоций).

СПИСОК ЛИТЕРАТУРЫ

1. Барулин А.Н. К аргументации полигенеза // Разумное поведение и язык. Вып. 1: Коммуникативные системы животных и язык человека. Проблема происхождения языка / Сост. А.Д. Кошелев, Т.В. Черниговская. М.: Языки славянских культур, 2008. С. 41–58.
2. Бурлак С.А. Происхождение языка: Факты, исследования, гипотезы. М.: Астрель: CORPUS, 2011.
3. Вартаков А.В. Антропоморфный метод распознавания эмоций в звучащей речи // Национальный психологический журнал. 2013. № 2. С. 64–73.
4. Вартаков А.В. Механизмы семантики: человек — нейрон — модель // Нейрокомпьютеры: разработка, применение. 2011. № 12. С. 54–64.
5. Вартаков А.В. Психофизиологический подход к проблеме распознавания речи // В сб. Труды второго междисциплинарного семинара “Анализ разговорной русской речи” (АРЗ–2008). СПб.: ГУАП, 2008. С. 82–93.
6. Вартаков А.В., Вартакова И.И. Эмоции, мотивация, потребность в филогенезе психики и мозга // Вест. Моск. ун-та. Сер. 14. Психология. 2005. № 3. С. 20–35.
7. Вартаков А.В., Косарева Ю.И. Эмоции человека и обезьян: субъективное шкалирование вокализаций // Вест. Моск. ун-та. Сер. 14 Психология. 2015. № 2. С. 93–109.
8. Вартаков А.В., Терещенко Л.В., Латаков А.В., Бурлак С.А. Звуковой язык эмоций человека и обезьян // Национальный психологический журнал. 2014. № 4. С. 81–94.
9. Виденеева Н.М., Хлудова О.О., Вартаков А.В. Эмоциональные характеристики звучащего слова // Журн. высш. нерв. деят. 2000. Т. 50. Вып. 1. С. 29–43.
10. Горелов И.Н., Седов К.Ф. Основы психолингвистики. М.: Лабиринт. 1997. С. 3–150.
11. Кедрова Г.Е., Вартаков А.В. Экспериментальное исследование акустических коррелятов кинакема в русском языке // В сб. Труды шестого междисциплинарного семинара “Анализ разговорной русской речи” (АРЗ–2012). Т. 3. СПб.:СПИИРАН, 2012. С. 4–8.

12. Патент RU 2268504 C9 (2006.01) Способ распознавания фонем речи и устройство для реализации способа. Автор(ы): Гиголо Людмила Антоновна (RU), Сахаров Виталий Олегович (RU) Патентообладатель(и): Открытое акционерное общество "Корпорация "Фазотрон – Научно-исследовательский институт радиостроения" (RU).
13. Патент RU 2296376 C2 (2006.01) Способ распознавания слов речи. Автор(ы): Гиголо Людмила Антоновна (RU), Сахаров Виталий Олегович (RU) Патентообладатель(и): Открытое акционерное общество "Корпорация "Фазотрон – Научно-исследовательский институт радиостроения" (RU).
14. Плоткин В.Я. Фонологические кванты. Новосибирск: ВО "Наука" Сибирская издательская фирма, 1993.
15. Сидоров К.В., Филатова Н.Н. Анализ признаков эмоционально окрашенной речи // Вестник Тверского государственного технического университета. Тверь, 2012. Вып. 20. С. 26–31.
16. 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. Вып. 18. Тверь, 2011. С. 81–85.
17. Симонов П.В. Эмоциональный мозг. М.: Наука, 1981.
18. Симонов П.В. Лекции о работе головного мозга: потребностно-информационная теория высшей нервной деятельности. М.: Наука, 2001.
19. Соколов Е.Н. Сферическая модель интеллектуальных операций // Психол. журн. 2001. Т. 22. № 3. С. 49–56.
20. Соколов Е.Н., Вайтнявичюс Г.Г. Нейроинтеллект: от нейрона к нейрокомпьютеру. М.: Наука, 1989.
21. Сорокин В.Н. Моторная теория восприятия речи и теория внутренней модели // Информационные процессы, 2007. Т. 7. № 1. С. 1–12.
22. Фант Г. Акустическая теория речеобразования. М.: Наука, 1964.
23. Чистович Л.А., Венцов А.В., Гранстрем М.П., и др. Физиология речи. Восприятие речи человеком. Л.: Наука, 1976.
24. Якобсон Р. Типологические исследования и их вклад в сравнительно-историческое языкознание // Новое в лингвистике. Вып. III. М., 1963. С. 95–105.
25. Якобсон Р., Халле М., Фант Г. Введение в анализ речи: различительные признаки и их корреляты // Новое в лингвистике. 1962. В. 2.
26. Bondarko L.V. The syllable structure of speech and distinctive features of phonemes. *Phonetica*, 1969. V. 20. P. 1–40.
27. Callan D.E., Callan A.M., Kroos Ch., Vatikiotis-Bateson E. Neural processes underlying perception of audio-visual speech production. *Proc. 5th Seminar on Speech Production, Kloster Seeon*. 2000. P. 273–276.
28. Carroll J.D., Chang J.J. Analyses of individual differences in multi-dimensional scaling via an N-way generalization of Eckart-Young decomposition // *Psychometrika*. 1970. V. 35. P. 283–319.
29. Jakobson R., Fant G., Halle M. Preliminaries to speech analysis. The distinctive features and their correlates // *Massachusetts Inst. of Technology, Acoustics Lab. Techn. Rep.*, 1952. № 13. P. 19–58.
30. Lindblom B. Role of articulation in speech perception: Clues from production // *JASA*. 1996. V. 99. № 3. P. 1683–1692.
31. Sekiyama K., Sugita Y. Auditory-visual speech perception examined by brain imaging and reaction time. *Proc. 7th Int. Conf. on Spoken Language Processing, Denver*. 2002. P. 1693–1696.
32. Torgerson W.S. *Theory and Methods of Scaling*. N.Y. Willey, 1958. P. 227–297.

SUBJECTIVE PERCEPTION AND FORMAL COMPUTER ANALYSIS OF RUSSIAN LANGUAGE'S SOUNDS

A. V. Vartanov*, V. V. Shvyrev**

** PhD, senior research officer, department of psychology, M.V. Lomonosov Moscow State University, Moscow;*

*** Student, department of psychology, M.V. Lomonosov Moscow State University, Moscow*

Subjective space of Russian language sounds' perception (46 phonemes, including variants of vowels and consonants, $N=17$) has been constructed by means of metric multidimensional scaling; its dimension is 15 (is composed of four conditionally primary and 11 conditionally secondary parameters). Formal (computer) analysis of the same sound patterns by means of previously proposed algorithm has revealed that only four factors that correspond to the first four subjective distinctive features (when averaging signal for the time of sounding) are enough for their description. The rest of 11 subjective distinctive features correspond to sub-phonemic quanta – kinakemas (extracted according to V.Ya. Plotkin). The revealed form of spectral representation of factors of set of sound signal-phonemes under study is similar to the same characteristics of other sets of sound signals that were used previously in the studies of emotions' manifestations in the sounding speech and were interpreted as certain emotional parameters in four-dimensional spherical model of emotions.

Key words: speech, speech sounds, subjective perception, multidimensional scaling, emotions.