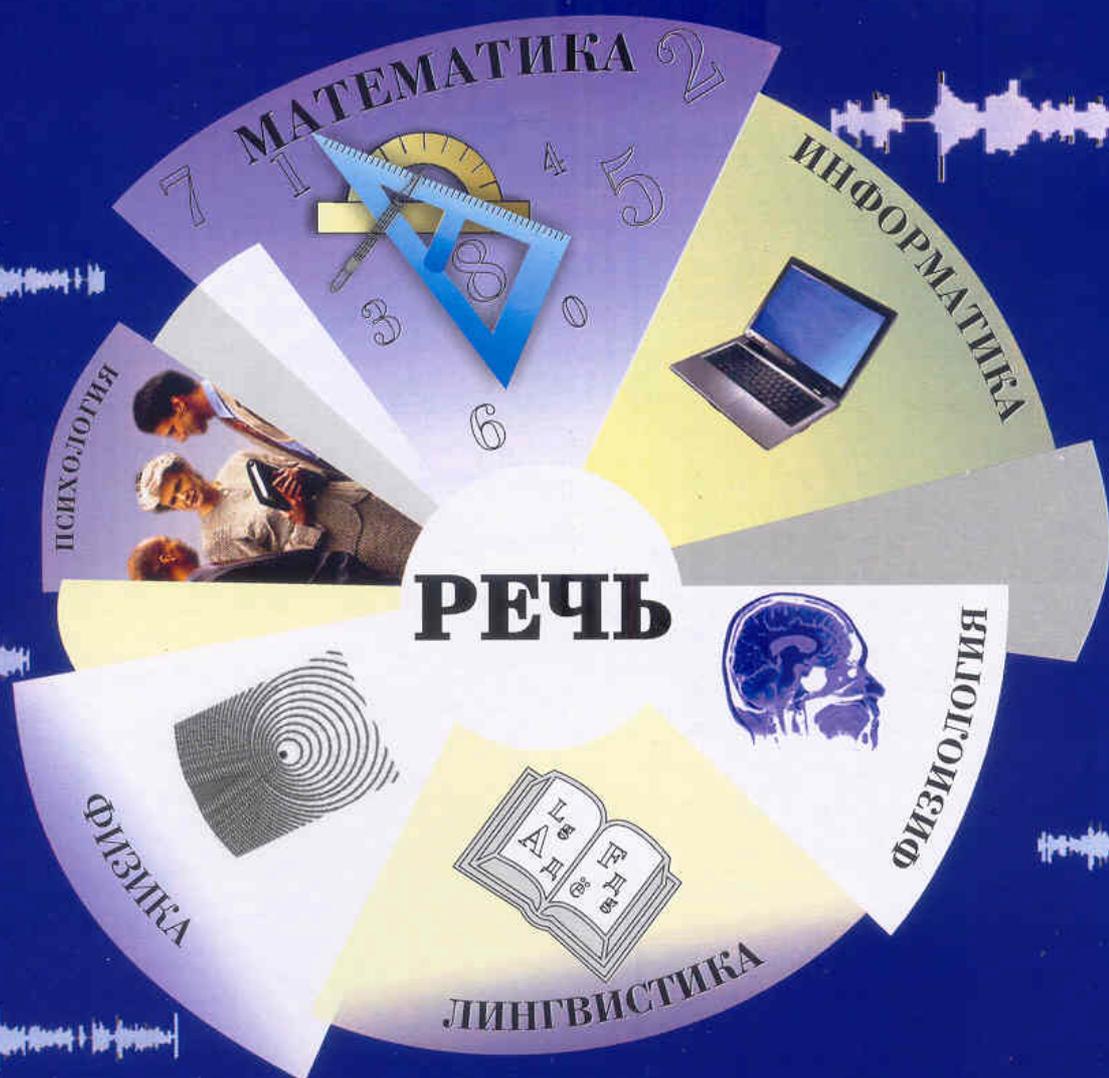


Второй
междисциплинарный семинар
«Анализ разговорной русской речи»

АР³ - 2008



27-28 августа 2008 года,
Санкт-Петербург, СПИИРАН

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Санкт-Петербургский институт информатики и автоматизации
Санкт-Петербургский государственный университет
аэрокосмического приборостроения

Второй
междисциплинарный семинар
«Анализ разговорной русской речи»

АР³ - 2008

27 – 28 августа 2008 года,
Санкт-Петербург, СПИИРАН

Санкт-Петербург
2008

УДК 004.522

РОССИЙСКАЯ АКАДЕМИЯ НАУК

Санкт-Петербургский институт информатики и автоматизации

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Санкт-Петербург, 199178, 14 линия, 39.

<http://www.spiiras.nw.ru/speech>

Труды второго междисциплинарного семинара
«Анализ разговорной русской речи» (АР³ - 2008). – СПб.: ГУАП, 2008. – 102 с.
ISBN 978-5-8088-0356-5

Издание представляет собой сборник докладов, сделанных на заседаниях второго междисциплинарного семинара «Анализ разговорной русской речи» (АР³ - 2008), проходившего 27 – 28 августа 2008 года в Санкт-Петербургском институте информатики и автоматизации Российской академии наук. Семинар посвящен обсуждению особенностей разговорной речи и возможных подходов к автоматическому анализу русской речи. Междисциплинарный подход к изучению речи позволит скорее продвинуться в моделировании речевой деятельности и решить фундаментальную проблему человеко-машинного диалога.

УДК 004.522

Статьи печатаются в авторской редакции.

Издание осуществлено за счет средств гранта РФФИ № 08-07-06030-г.

ISBN 978-5-8088-0356-5

© СПИИРАН, 2008

© Коллектив авторов, 2008

© ГУАП, 2008

Оглавление

<i>С.В. Андреева.</i> Универсальные основания выделения структурных единиц дискурса	4
<i>А.А. Кибрик, О.А. Савельева-Трофимова.</i> Субклаузальные дискурсивные единицы в устной речи	10
<i>Н.В. Сердобольская, А.С. Шаврыгина.</i> Позиция союза в обстоятельственных придаточных предложениях устной спонтанной речи	14
<i>Н.М. Шитова.</i> Относительное предложение в кодифицированном литературном языке и в разговорной речи.....	19
<i>М.С. Шматова.</i> Образование отглагольных существительных на <i>-ка</i> от глаголов несовершенного вида в русской устной речи	25
<i>И.Е. Апушкина, А.В. Венцов, Н.А. Слепокурова.</i> О фонетическом слове.....	31
<i>Г.В. Дорохина, Т.В. Ермоленко.</i> Модели пофонемного распознавания речи, явно использующие свойства языка и речи	37
<i>П.А. Скредин, В.В. Евдокимова.</i> Вариативность реализаций гласных фонем в спонтанной речи и чтении	42
<i>С.О. Тананайко, Л.А. Васильева.</i> Фонологическое и фонетическое описание гласных в первообразных союзах	48
<i>Б.М. Лобанов, Л.И. Цирульник.</i> Автоматизация анализа просодических характеристик для экспериментальных исследований и синтеза речи по тексту.....	54
<i>А.А. Конев, Р.В. Мещеряков.</i> Оценка вокализованных участков речевого сигнала	59
<i>А.В. Венцов.</i> Спонтанная речь и ментальный лексикон	66
<i>И.С. Кипяткова.</i> Обзор подходов к моделированию спонтанной речи	70
<i>А.Б. Викторов, М.В. Ескевич.</i> Адаптируемое под язык акустическое моделирование для распознавания речи.....	78
<i>А.В. Вартанов.</i> Психофизиологический подход к проблеме распознавания речи.....	82
<i>Ал.Б. Леонтьева, Р. Йекель, И.А. Кагиров, А.Л. Ронжин.</i> Анализ речи и поведения студентов при компьютерном тестировании знаний в многомодальном режиме	94

Универсальные основания выделения структурных единиц дискурса

*Саратовский государственный университет,
г. Саратов, Россия
svandreeva@rambler.ru*

Основания и принципы выделения структурных единиц устной речи были обоснованы в наших работах [Андреева 2004, 2006]. Сама типология структурных единиц устной речи на основе их строения и функций представлена в том числе и в материалах первого междисциплинарного семинара «Анализ разговорной русской речи» [Андреева 2007]. Цель данной статьи — раскрыть принципы, которые могут дать основания для анализа любой формы речи (как устной, так и письменной).

До определенного времени традиционная лингвистика была ориентирована преимущественно на описание статической модели языковой системы. Грамматические понятия складывались под влиянием письменной речи. Со второй половины XX в. внимание лингвистов сосредоточилось на «человеке говорящем» и язык как знаковая система перестал быть в центре внимания исследователей. По свидетельству ученых, «запоздалый интерес» к устной форме речи возник по-настоящему лишь в 60-е годы и походил на всеобщую «золотую лихорадку» [Костомаров 2005: 137].

В нескольких российских научных центрах (под руководством Е.А. Земской, О.А. Лаптевой, О.Б. Сиротининой), а также за рубежом активно и масштабно начала изучаться разговорная речь. Лингвисты саратовской школы, в частности, выявляли сущностные признаки и особенности русской разговорной речи [Сиротинина 1974], исследовали свойства интонационной организации спонтанной речи [Полищук, Игнатов, Сиротинина, Штромбергер 1971; Андреева, Полищук, Сокиркина 1984], рассматривали устную литературную речь в онтогенезе [Седов 1986], анализировали разговорную речь в сопоставлении с другими функциональными стилями современного русского литературного языка [Разговорная речь в системе... 1983, 1992] и мн. др.

В результате коллективных усилий «разговорников» была подтверждена идея о существовании лингвистики речи наряду с лингвистикой языка. Вместе с тем во весь рост встали проблемы, затрудняющие создание лингвистики речи как дисциплины, призванной раскрыть закономерности построения и функционирования речевых единиц.

Ученые пришли к заключению о несовпадении границ языковых и речевых единиц, и начался поиск структурных единиц устной речи. В результате исследований петербургских ученых [см.: Богданова 2007], было выяснено, что подход к поиску границ предложений через интонационное членение спонтанных текстов не дает удовлетворительных результатов, т.к. интонационное и синтаксическое членение устной речи соответствуют друг другу не более чем в 54 % случаев [Степихов 2005]. Одной из попыток поиска конструктивной единицы была идея о структурно-синтаксическом единстве, под которым понимается предикативный центр с зависимыми словоформами. Однако, по справедливому замечанию Н.В. Богдановой, «при том, что это весьма удачный способ обойти основные трудности синтаксического членения спонтанного текста и удобная единица описания его структуры, многих сложностей интерпретации такой подход не снимает. На пути создания грамматики речи вопросов... пока еще значительно больше, чем ответов. Поиск единиц описания спонтанной речи (метапонятий речи) должен быть продолжен» [Богданова 2007: 193].

Ученые столкнулись с многочисленными затруднениями и в синтаксической квалификации речевых фактов. В связи с этим в научной литературе стала обсуждаться проблема недостаточности выделения только одной коммуникативной единицы — предложения. Появляется понятие функциональных заместителей предложения, т.е. специфических коммуника-

тивных единиц, не соответствующих грамматической модели предложения, выполняющих речерегулирующие функции, выражающих большой диапазон модальных значений (*Еще чего! Ни за что! Да ну!* и др.). Терминологическая пестрота, наблюдавшаяся в обозначении этого речевого феномена, отражает, на наш взгляд, не только неопределенность его статуса и в связи с этим разные подходы ученых, но и диффузный характер единиц данного класса: «слова-предложения» (Грамматика русского языка 1954), «релятив» (Г.В. Валимова, О.Б. Сиротинина), «метакоммуникативные сигналы слушающего» (Т.Д. Чхетиани), «фразоиды» (Л. Теньер), «сентенсоиды» (И.И. Прибыток), «коммуникемы» (В.Ю. Меликян) и др.

В последнее время общепринятым становится термин «коммуникатив» (В.Ф. Киприянов, В.Н. Шаронов, Е.Ю. Викторова, Т.Н. Колокольцева, Г.Ф. Гаврилова, Е.К. Кожина и др.). Мы, вслед за другими сторонниками синтаксического подхода к такого рода единицам, считаем, что доминирующим признаком является их функциональное своеобразие — речерегулирующая роль. Термин «коммуникатив» отражает как раз функциональную сторону явления, т.е. возможность использоваться в речи наряду с другими единицами коммуникации.

Таким образом, несмотря на большое количество работ по устной речи (разговорной, научной, телевизионной и т.д.), система синтаксических единиц еще не построена, а та система, что была разработана для письменной речи, для устной речи (и современной письменной) не годится. Проведенный нами анализ большого фактического материала устной речи показал, что неоднозначные с точки зрения существующих классификаций речевые факты (не-предложения и не-коммуникативы в их привычном понимании) составляют — в зависимости от типа дискурса — от 2,5% до 13%. Следовательно, проблема «инвентаря тех единиц, которые должны стать стержнем любой грамматики, в том числе и грамматики речи» [Богданова 2007: 190] продолжает оставаться первостепенной.

Представленная в наших работах типология структурных единиц устной речи базируется на шести принципах. Рассмотрим их.

1. Только двусторонний парадигматико-синтагматический подход продуктивен для описания континуума речи.

В современной лингвистике наблюдается выраженная тенденция перехода от исследования фрагментов системы языка к исследованию системы «языка–речи» в целом. Важнейшей становится проблема интеграции вычлененных учеными языковых единиц и уровней лингвистического анализа. В фокусе внимания — двусущностное единство «язык–речь» как функционирующее устройство во взаимодействии всех его составляющих.

Мы исходим из того, что любая коммуникативная единица в конкретном речевом акте синтезирует в себе языковое и речевое. В основе высказывания лежит диалектическое взаимодействие: одной стороной высказывание обращено к языку (синтаксическая форма), а другой — к речи (непосредственная реализация модели предложения). Поскольку речевое и языковое существуют в органическом единстве, можно говорить о двуплановости структурных единиц дискурса: в них происходит соединение (слияние) конструктивно-синтаксических (это элементы синтаксического уровня языка) и коммуникативно-функциональных (это компоненты коммуникативного уровня речи) характеристик. Двуплановостью этих единиц обусловлен предпринятый нами двусторонний анализ: конструктивно-синтаксический (парадигматический) и синтагматический (с учетом контекстуальных и речевых факторов).

В основу нашей типологии положены следующие ведущие дифференциальные признаки: наличие/отсутствие коммуникативной самодостаточности единицы, особенности реализации в ней категории предикативности, соотношение диктумной (пропозициональной) и модусной составляющих, соответствие грамматической модели, характер передаваемой информации, функциональная направленность. Учет данных признаков обеспечивает комплексное рассмотрение языковой природы и речевых характеристик структурных единиц дискурса.

2. Объединение классификационных признаков и системного взаимодействия единиц.

Классификационный подход предполагает, прежде всего, определение дифференциальных признаков и на их основе разграничение структурных единиц. Системный подход нацелен на определение интегральных признаков и выявление взаимосвязей между единицами. Сочетание классификационного и системного подходов дает возможность не только вычлнить структурные единицы дискурса, но и раскрыть их внутренние связи.

3. Учет разного вида информации, передаваемой единицами дискурса.

В современных лингвистических исследованиях разрабатывается идея о двух уровнях речевого общения: коммуникативно-информативном (основном) и коммуникативно-организующем (вспомогательном) [Менг 1982; Пономарева 1991; Безяева 2002 и др.]. Как установили исследователи, со стороны адресата коммуникация предполагает не только декодирование семантики языковых знаков, но и раскрытие использованных глубинных смыслов текста, пресуппозиций, авторских стратегий, учет фоновых знаний, ситуации общения и структуры дискурса. К этим выводам близки теоретические положения О.Б. Йокоямы о семи видах знания, в том числе метаинформационном, совершенно необходимом для успешного осуществления информационного дискурса [Йокояма 2005]. Для нас важно утверждение исследовательницы о том, что средства передачи метаинформационного знания (знания кода и дискурсивной ситуации) вырабатываются в каждом языке, следовательно, само это знание универсально для разных языковых систем и без него невозможен прием фактуальной информации.

Передаваемую в речевом общении информацию первоначально (при общем подходе) можно подразделить на 2 вида: фактуальную и коммуникативную. В дальнейшем, как показало наше исследование, оказалась необходимой детализация, дифференциация разновидностей коммуникативной информации. Были выделены ее составляющие: собственно коммуникативная информация, отражающая взаимодействие между собеседниками, направленная на организацию общения; и такие факультативные (они могут и отсутствовать) виды информации, как дискурсивная, направленная на организацию дискурса и ориентацию в нем адресата, а также сигнальная «информация» о психо-физиологическом или эмоционально-чувственном состоянии говорящего.

В соответствии с этим в континууме коммуникации были выделены три зоны речевых единиц: коммуникативно-информативная зона (передает фактуальную информацию), коммуникативно-регулятивная (передает коммуникативную информацию), дискурсивно-структурирующая (передает дискурсивную информацию). Кроме того, в континууме речевой коммуникации существует, хотя и стоит особняком и реализуется далеко не во всех типах устного дискурса, зона безусловно-рефлекторных проявлений (несущих сигнальную «информацию»). В результате взаимодействия выделенных зон образуются пространства их наложения.

Таким образом, классифицируя структурные единицы дискурса, мы опирались на функциональную направленность единицы, разграничивающую составные части высказывания и не позволяющую объединять их в единое элементарное синтаксическое целое.

4. Учет прерывно-непрерывной природы единства «язык–речь».

Думается, что проблема языкового континуума, диалектики непрерывного и прерывного в языке–речи является одной из важнейших в лингвистике. Сложность типологии структурных единиц дискурса обусловлена не только тем, что в динамике речи первичная языковая основа может изменять свои потенциальные свойства и, соответственно, функцию, но и тем, что система структурных единиц дискурса характеризуется наличием переходных явлений, гибридных образований. Среди структурных единиц дискурса можно выделить информативно однозначные и информативно синкретичные.

Синкретичные единицы представляют особый интерес, поскольку на них мало обращалось внимания другими исследователями. Например, типичные предикативные единицы могут синкретично контаминировать фактуальную и коммуникативную информацию. В речи

телеведущих, например, они используются в качестве конвенционализованной формулы прощания (коммуникативный регулятор) и информирования о следующем выходе в эфир: *Завтра в это же время дня обязательно увидимся!*. Синкретичные единицы относятся к пространствам наложения зон речевой коммуникации, разным звеньям шкалы переходности [подробнее см.: Андреева 2006: 110].

Прерывно-непрерывная природа единства «язык–речь» проявляется и в том, что практически по всем названным выше дифференциальным признакам выстраивается шкала: между типичной предикативной единицей и коммуникативом располагаются единицы с различной степенью проявления рассматриваемого признака. Приведем пример шкалы по признаку «реализация категории предикативности»: реальная предикативность, далее — актуально-ситуативная предикативность (при распространении на единицу предикативных характеристик, заданных ситуативно), еще дальше — формальная предикативность, наконец, — отсутствие предикативности в структурах непредложенческих (релятивных).

Выявление информативно однозначных и синкретичных единиц позволило осознать, что комплекс зон имеет признаки полевой системности, диффузность зон определяет и диффузность функций данных единиц. Это понимание дало возможность графически представить динамичность полевой структуры зон речевой коммуникации [см.: Андреева 2006: 109].

5. В результате конкретного анализа разных дискурсов устной речи определен еще один принцип — необходимость разграничения основных и вспомогательных структурных единиц.

Основные единицы речевой коммуникации являются строительным материалом всех исследованных устных дискурсов (без них не обходится коммуникация). Основные единицы речевой коммуникации можно определить следующим образом: это наименьшие структурные единицы как предложенческого типа на основе свободного конструирования, характеризующиеся реальной, формальной или актуально-ситуативной предикативностью, так и клишированные единицы релятивного типа, не обладающие признаками грамматической моделируемости и предикативности, но отличающиеся семантической и интонационной завершенностью и вместе с предложенческими единицами образующие «костяк» («тело») дискурса, особенно в устной диалогической речи.

Наряду с основными единицами в речевой коммуникации функционируют и вспомогательные: диалогические повторы, дискурсивы и речевые жесты. Они не составляют синтаксический «костяк» дискурса, т.к. отличаются от основных единиц своей факультативностью. Следовательно, они не «конструируют», а только организуют дискурс, не имея самостоятельного значения в нем.

В последнее время дискурсивные слова и близкие к ним единицы стали объектом пристального внимания исследователей-русистов [Николаева 1985; Кобозева 1991; Баранов и др. 1993; Словарь структурных слов 1997; Дискурсивные слова... 1998; Дискурсивные слова... 2003; Сиротина 2005, 2008; Андреева 2006 и др.].

В англистике дискурсивные единицы («discourse markers», «inserts», «function words») получают освещение в современных грамматиках [Longman Grammar... 2000; A Comprehensive Grammar... 2003]. Англоязычная научная литература отражает интерес к самым разнообразным дискурсивным элементам: в фокусе внимания ученых находятся связки («connectives») [Rey 1997], дискурсивные элементы («discourse markers», «formulae») как монолога, так и диалога [Schiffrin 1988; Kim, Glass, Freedman, Evans 2000; Louwarse, Mitchell 2003].

В роли собственно дискурсивов, актуализаторов, заполнителей пауз хезитации и многих других «как бы лишних» слов выступают прежде всего частицы, предложно-падежные формы, служебные и полуслужебные лексемы, а также модальные и некоторые знаменательные слова (*только, лишь, всего-навсего, по всей видимости, просто, разумеется, скажем* и др.). Эти так называемые собственно-речевые наслоения, частично или полностью лишённые номинативного содержания (предметно-фактической информативности), обеспечивают надёжность в передаче информации, однозначность её восприятия, лёгкость произношения.

Разграничение основных и вспомогательных структурных единиц дискурса чрезвычайно важно для автоматического анализа разговорной речи. Это актуально, в частности, для разработки «специальных моделей мусора, обученных на распознавание артефактов речи, заполненных пауз и других элементов, характерных для спонтанной речи и серьёзно ухудшающих качество работы системы. Своевременная локализация таких помех и удаление их из последующей обработки позволит существенно повысить производительность работы системы» [Ронжин 2007: 12].

6. Сочетание дедуктивного и индуктивного подходов.

При анализе основных единиц мы опирались как на дедуктивный подход (шли от гипотезы, что они информативны), так и на индуктивный (отсутствие или наличие информации другого вида наряду с фактуальной в каждой единице конкретного дискурса). Поэтому члены предложения в качестве самостоятельных речевых единиц нами не вычленились, парцелляты считались неполными единицами, а речевые жесты, разные виды дискурсивов вычленились в качестве особых вспомогательных речевых единиц, поскольку они не входят в диктум предложения, а следовательно, не видоизменяют фактуальную информацию, привнося информацию другого вида. При анализе вспомогательных единиц речевой коммуникации применялся чисто индуктивный подход. Как показал индуктивный путь, в речи возможно ослабление информативности единиц, вплоть до полной ее утраты. При вычленении вспомогательных единиц учитывались также следующие моменты: 1) коммуникативное, интонационное их выделение и образование ими особой синтагмы, 2) возможность при определенных условиях быть самостоятельной единицей.

Таким образом, разработанная на основе рассмотренных принципов типология единиц речевой коммуникации (детализация существующих классификаций) позволила описать наименьшие структурные единицы устной речи, определить их системное взаимодействие и коммуникативный потенциал, тем самым способствуя решению актуальной для русистики проблемы. Эта типология речевых единиц служит опознаванию «неопознанных синтаксических объектов» реальной устной речи. Совокупность полученных результатов создает фундамент нового направления синтаксиса — изучения не только отдельных синтаксических единиц, а всей реальной коммуникации как динамично меняющегося полевого устройства.

На новом этапе исследования была осуществлена попытка приложения этой классификации синтаксических единиц не только к устной, но и к письменной речи. Оказалось, что она работает и на материале письменной речи. Следует отметить, что даже коммуникативы, функционирующие в основном в устной речи (или в ее стилизации), возможны в сугубо письменной речи (например, при заполнении анкет — использование наиболее типичных коммуникативов *да/нет*). Звуковые жесты фактически наблюдаются только при стилизации в письменной речи устных дискурсов.

Задачей дальнейших исследований является выяснение частотности, регулярности, особенностей функционирования всех выделенных в процессе работы единиц — с преобладанием тех или иных в разных типах и сферах как устного общения (непосредственного и опосредованного различными техническими устройствами), так и письменного.

Литература

1. Андреева С.В. Типология конструктивно-синтаксических единиц в русской речи // Вопросы языкознания. 2004. №5.
2. Андреева С.В. Речевые единицы русской речи: Система, зоны употребления, функции. Изд. 2-е, испр. М: УРСС, 2006.
3. Андреева С.В. Разговорная русская речь и ее единицы // Труды первого междисциплинарного семинара «Анализ разговорной русской речи» (АРЗ-2007). СПб.: ГУАП, 2007.
4. Андреева С.В., Полищук Г.Г., Сокиркина Л.И. Анализ восприятия смыслоразличительных особенностей интонации // Экспериментально-фонетический анализ речи. Л., 1984.
5. Баранов А.Н., Плунгян В.А., Рахилина Е.В. Путеводитель по дискурсивным словам русского языка. М., 1993.

6. Дискурсивные слова русского языка: опыт контекстно-семантического описания. М., 1998.
7. Дискурсивные слова русского языка: контекстное варьирование и семантическое единство. М., 2003.
8. Безяева М.Г. Семантика коммуникативного уровня звучащего языка: Волеизъявление и выражение желания говорящего в русском диалоге. М., 2002.
9. Богданова Н.В. Метапонятия языка и речи (к вопросу о поиске единиц описания живой речи) // Инновации в исследованиях русского языка, литературы и культуры. Т. 1. Пловдив, 2007.
10. Ёкояма О. Когнитивная модель дискурса и русский порядок слов. М., 2005.
11. Кобозева И.М. Проблемы описания частиц в исследованиях 80-х годов // Прагматика и семантика: Сб. науч.-аналит. обзоров. М., 1991.
12. Костомаров В.Г. Наш язык в действии: Очерки современной русской стилистики. М., 2005.
13. Менг К. Проблема анализа диалогического общения // Текст как психолингвистическая реальность. М., 1982.
14. Николаева Т.М. Функции частиц в высказывании (на материале славянских языков). М., 1985.
15. Полищук Г.Г., Игнатов Б.Ф., Сиротинина О.Б., Штромбергер Л.В. О роли различных компонентов интонации в передаче синтаксической информации // Развитие фонетики современного русского языка. М., 1971.
16. Пономарева Т.В. Сочетаемость информативных высказываний разных прагматических типов с этикетными контактными стимулами // Языковые единицы. Л., 1991.
17. Разговорная речь в системе функциональных стилей современного русского литературного языка / Под ред. О.Б. Сиротининой. Лексика. Саратов, 1983.
18. Разговорная речь в системе функциональных стилей современного русского литературного языка / Под ред. О.Б. Сиротининой. Грамматика. Саратов, 1992.
19. Ронжин А.Л. Особенности автоматического распознавания разговорной русской речи // Труды первого междисциплинарного семинара «Анализ разговорной русской речи» (АРЗ-2007). СПб.: ГУАП, 2007.
20. Седов К.Ф. Устная литературная речь в онтогенезе // Функционирование языка в разных видах речи. Саратов, 1986.
21. Сиротинина О.Б. О синтаксическом статусе некоторых компонентов дискурса // *Gameni și idei. Studii de filologie. Cluj. Napoca: Editura Risoprint, 2005.*
22. Сиротинина О.Б. Дискурсивные слова и их отношение к пунктуационной системе русского языка // Исследования по семантике: Межвуз. науч. сб. Вып. 24. Уфа, 2008.
23. Словарь структурных слов русского языка / Под ред. В.В. Морковкина. М., 1997.
24. Степихов А.А. Соотношение синтаксического и интонационного членения в спонтанном монологе: Дис... канд. филол. наук. СПб., 2005.
25. *A Comprehensive Grammar of the English Language / Quirk R., Greenbaum S., Leech G., Svartvik J. Harlow, 2003.*
26. Kim J.H., Glass M., Freedman R., Evans M.W. Learning the Use of Discourse Markers in Tutorial Dialogue for an Intelligent Tutoring System // *Procc. 22nd Annual Meeting of the Cognitive Science Society Cog Sc '2000. Philadelphia, 2000.*
27. *Longman Grammar of Spoken and Written English / Biber D., Johansson S., Conrad S., Finegan E. Harlow, 2000.*
28. Louwense M.M., Mitchell H.H. Toward a Taxonomy of a Set of Discourse Markers in Dialog: A Theoretical and Computational Linguistic Account // *Discourse Processes. № 35 (3). 2003.*
29. Rey J. Discourse Markers: a Challenge for Natural Language Processing // *AI Communications. 1997. № 10.*
30. Schiffrin D. *Discourse markers. Cambridge University Press, 1988.*

А.А. Кибрик,* О.А. Савельева-Трофимова**

Субклаузные дискурсивные единицы в устной речи (на материале корпуса «Рассказы о сновидениях»)¹

*Институт языкознания РАН, ** МГУ им. М.В. Ломоносова
г. Москва, Россия

* kibrik@comtv.ru, ** savel.trol@gmail.com

1. Введение

Данное исследование основано на корпусе «Рассказы о сновидениях», подготавливаемом в настоящее время к публикации под руководством А.А. Кибрика и В.И. Подлесской. Это корпус устных русских рассказов детей и подростков о сновидениях. Корпус представляет собой транскрипт примерно 2 часов устной разговорной речи и является достаточно представительным по отношению к этому типу дискурса. В транскрипте речь делится на элементарные дискурсивные единицы (ЭДЕ), для каждой из которых маркированы основные просодические характеристики. Общий объем корпуса составляет 3729 ЭДЕ. Как известно, в устном дискурсе большинство ЭДЕ представляет собой клаузные структуры. Так же обстоит дело и в нашем корпусе. Данное исследование посвящено существенному меньшинству ЭДЕ, которые передают пропозициональную информацию, но при этом по объему «недоотягивают» до целой клаузы.

2. Препозитивные и постпозитивные ЭДЕ

Данный доклад посвящен препозитивным (L-) (1) и постпозитивным (R-) (2), (3) непредикативным ЭДЕ, по просодическим критериям вынесенным за пределы клаузы, к которой они семантически относятся (далее *опорной ЭДЕ*):

(1) Z33

1,7	3	..(0.4) /Зайчик вот,
2,9	4	...(0.5) он был в \лесу.

(2) N01

11,3	7	и /я была в таких \жёлтых таких /туфлях таких,
13,7	8	..(0.2) с-с такими с \шнурочка↑ми,

(3) Z08

0,6	1	Бегала бездомная /собака,
2,2	2	..(0.2) \большой такой \чёрный \дог.

Целью работы является описание R- и L-ЭДЕ в Корпусе, а также выявление их основных характеристик, функций и контекстов появления. Рассматриваемые единицы очень характерны для разговорной речи, в литературе известны как именная тема или обособленные уточнения. Однако их описания, полностью основанного на корпусных данных русского языка, пока нет. Мы рассматриваем указанные единицы по классам, выделяемым на основании:

- их месторасположения относительно опорной клаузы (L- vs. R-ЭДЕ);
- семантико-синтаксической структуры относительно опорной клаузы (эхо vs. парцелляция/приращение);

¹ Исследование проведено при поддержке гранта РФФИ №06-06-80470.

с) когнитивным свойствам запланированности/незапланированности (парцелляция vs. приращение).

L- и R-ЭДЕ составляют 10,9% ЭДЕ всего Корпуса. Общее количество R-ЭДЕ значительно превосходит L-ЭДЕ (см. Рис. 1).

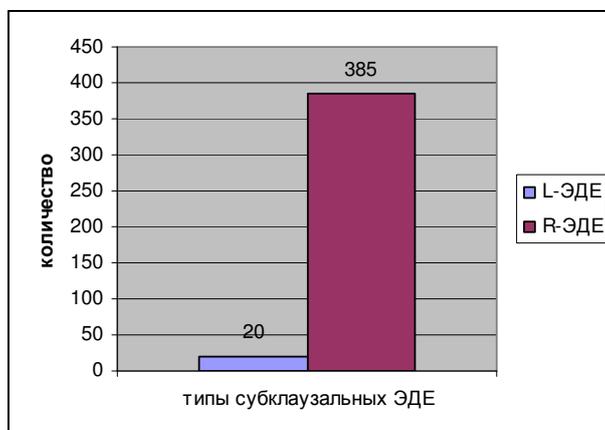


Рис. 1. Количественные данные по типам субклаузалных ЭДЕ

Использование субклаузалных ЭДЕ — это нормальная характеристика речи всех говорящих. Частота употребления L- и R-ЭДЕ в речи — это личная стратегия говорящего, зависимостей от пола, возраста и др. подобных факторов нет. Наши подсчеты показали, что процент этих ЭДЕ в речи определенного говорящего постоянен.

3. Препозитивные субклаузалные ЭДЕ (топики)

Для L-ЭДЕ (топиков; см. пример (1)) мы предлагаем правила выделения, основанные на требовании наличия кореферентного члена в последующей ЭДЕ и непредикативности рассматриваемого фрагмента. По способу выражения топики представляют собой полные именные группы, референт топика в 70% случаев одушевленный и в 75% — агентивный. С когнитивной точки зрения топики противопоставлены по запланированности/ незапланированности появления. Были выявлены следующие тенденции:

- для запланированных топиков (55% от всех топиков) кореферентное местоимение играет роль агенса, количество конкурирующих референтов больше двух;
- для спорных случаев (25%) конкурирующих референтов нет, а в дальнейшем дискурсе дольше присутствуют запланированные;
- для незапланированных (20%) никаких предпочтений не выделяется.

При сопоставлении топиков с внутренними начальными темами было выявлено, что референту топика свойственно иметь 3-4 потенциальных конкурента. Всего 65% топиков имеют от 3 до 5 потенциальных конкурентов. С функциональной точки зрения, запланированный топик с данным референтом служит целям переключения референции, он появляется в контексте выбора из множества при наличии конкурирующих референтов. Запланированный топик с доступным референтом, по нашему предположению, служит целям актуализации, эшелонирования информации. Незапланированный топик появляется, когда нужно «спасти» начатую структуру.

4. Постпозитивные субклаузалные ЭДЕ: эхо

Самый многочисленный тип из рассматриваемых субклаузалных составляющих дискурса — это эхо (пример (3) выше). В эхо-ЭДЕ копируются морфо-синтаксические характеристики той составляющей опорной ЭДЕ, которую оно уточняет.

Эхо-ЭДЕ составляют 4,3% от всех ЭДЕ Корпуса. Почти 80% случаев эха — это именные и предложные группы. Эхо чаще всего уточняет полные (лексические) именные и пред-

ложные группы (55%), а также различные местоимения (25%). Чаще всего эхо относится к актанту (30% от общего количества эхо-ЭДЕ), сирконстанту (23,1%) или субъекту именной клаузы (17,5%).

Основные функциональные типы эха появляются, когда:

- a) уже введенному референту приписываются новые характеристики или свойства (40,6%);
- b) происходит уточнение местоимения опорной ЭДЕ (13,8%).
- c) имеет место противопоставление или контраст (13,1%);
- d) происходит смещение значения референциального выражения, приписывание новых свойств референту за счет изменения или частичного аннулирования уже названных (10%).

Мы рассмотрели эхо-ЭДЕ с коммуникативной точки зрения. Коммуникативный статус эхо-ЭДЕ в общем случае считался совпадающим со статусом референциально уточняемой составляющей опорной ЭДЕ. Мы получили, что 82% случаев эха — это рематические составляющие, остальные — тематические. Тематические эхо-ЭДЕ, уточняющие местоимения, мы предложили сопоставить с явлением, известным в литературе под термином антитопик. Анализ антитопика (пример (4) ниже) показал, что его появление связано с референцией, которая квалифицируется говорящим как неудачная или неоднозначная. Интересно, что выявленные нами антитопики отличаются от антитопиков К. Ламбрехта наличием на них движения тона в акценте, а также способностью разрывать основную клаузу, становясь сразу за уточняемым кореферентным местоимением.

(4) N69

59,4	35	..(0.3) что-о ..(0.2) 'она /думает,
61,2	36	что /она как бы \волше <u>б</u> ница,
62,4	37	<u>эта</u> \стару <u>ш</u> ка,

5. Постпозитивные субклаузальные ЭДЕ: парцелляция и приращение

Явления парцелляции и приращения в сумме составляют 5,8% от всех ЭДЕ Корпуса. Эти типы субклаузальных ЭДЕ, которые синтаксически могли бы входить в одну ЭДЕ с опорной, но на основании просодической автономности оформления они вынесены за ее пределы. Мы предложили правила различения этих явлений по просодическим (длина паузы), когнитивным (наличие хезитации) и семантико-синтаксическим критериям исходя из предположения, что они коррелируют с выражением когнитивной запланированности/незапланированности R-ЭДЕ.

Основной способ кодирования парцелляций и приращений — это именные и предлож-ные группы (чуть больше 75% для парцелляций, около 70% для приращений). Интересно, что среди парцелляций в 2 раза больше подлежащих, чем среди приращений. В отличие от эхо-ЭДЕ среди парцелляций/приращений мы не встречаем сегментированных актантов. Функционально парцелляция (пример (5)) — это средство эшелонирования информации, средство привлечения внимания. Приращение (пример (6)) служит говорящему в ситуации, когда он по завершении клаузы понимает, что забыл упомянуть релевантную для нее деталь, но может добавить эту деталь апостериорно, как бы пристроив к предыдущей клаузе. Функция приращения, таким образом, прежде всего состоит в экономии усилий говорящего — ему не нужно повторять уже построенную клаузу, и в то же время он может выразить то, что считает необходимым, понятным для слушающего образом.

(5) Z55

45,7	26	... (1.0) 'и-ии <u>вот</u> с /не-еба-а ... (0.7) ну \показался какой- <u>то</u> /высветился /ч-челове-ек бо-ольшой,
51,7	27	<u>с</u> /огромной –дуби-инкой,

(6) N15

0,2 1 –Мама чего-то попросила за чем-то \сходить.

2,4 2 ...(0.7) На \кухню.

Приведем количественные данные по типам R-ЭДЕ, см. Рис. 2.

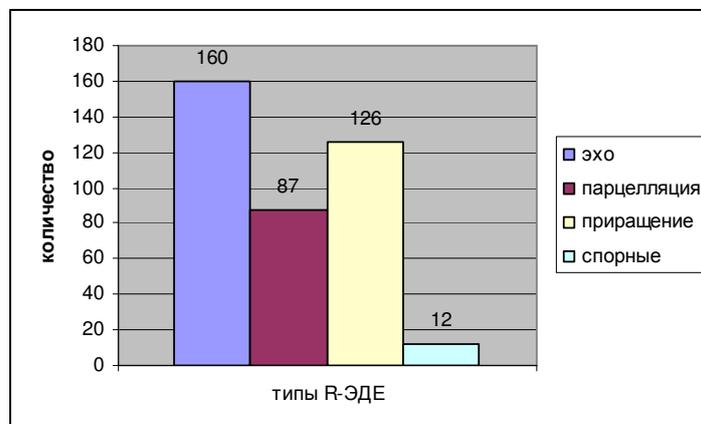


Рис. 2. Количественные данные по типам R-ЭДЕ

6. Заключение

Широкое использование субклаузалых ЭДЕ в корпусе «от противного» подтверждает основной принцип квантования устного дискурса: говорящий стремится организовать просодические единицы в виде клауз. Иногда по разным причинам ему это не удастся, и тогда он прибегает к различным несложным построениям, которые не являются клаузальными по своей структуре, но ориентированы на ту или иную опорную клаузу. В данной работе представлены количественные данные по корпусу, формулируются семантические и когнитивные причины, ведущие к образованию субклаузалых ЭДЕ. Признание важной роли субклаузалых ЭДЕ — это неременное условие для создания реалистичных и адекватных корпусов устной речи.

Литература

1. Земская Е.А. Русская разговорная речь: лингвистический анализ и проблемы обучения. М.: Русский язык, 1979.
2. Кибрик А.А., Подлеская В.И. *Проблема сегментации устного дискурса и когнитивная система говорящего* // Когнитивные исследования, вып. I. М.: ИП РАН, 2006.
3. Лаптева О.А. Русский разговорный синтаксис. М.: Наука, 1976.
4. Сиротинина О.Б. (ред.) Разговорная речь в системе функциональных стилей современного русского литературного языка. Синтаксис. М.: УРСС, 2003, с. 150-276.
5. Чейф У. *Данное, контрастивность, определенность, подлежащее, топика и точка зрения* // Новое в зарубежной лингвистике, вып. XI. М.: Прогресс, 1982, с. 277-316.
6. Givón T., Talmy L. *Coherence in text, coherence in mind* // Pragmatics and cognition, vol. 1(2), John Benjamin's, 1993, pp.171-227.
7. Lambrecht K. *Information Structure and Sentence Form: Topic, Focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press, 1994.
8. Lambrecht K. *Dislocation* // Language Typology and Language Universals, vol.2, Berlin-New York, 2001, pp. 1050-1078.

Н.В. Сердобольская,^{*} А.С. Шаврыгина^{**}

Позиция союза в обстоятельственных придаточных предложениях устной спонтанной речи

Московский городской психолого-педагогический университет,^{*}

г. Москва, Россия

^{*}serdobolskaya@gmail.com, ^{**}shavrygina@gmail.com

1. Постановка проблемы

В последние годы растет интерес к устной спонтанной речи. Хорошо известно, что устная спонтанная речь обнаруживает целый ряд морфологических, синтаксических и лексических особенностей, в той или иной степени нарушающих нормы письменного языка (см. [Лаптева 1976]; [Земская 1987]; [Китайгородская, Розанова 1999]). Однако зачастую основное внимание в различных исследованиях уделяется морфологии и лексической семантике русской разговорной речи (далее: РР), тогда как особенности синтаксиса РР остаются за пределами внимания исследователей.

В настоящем докладе рассматриваются проблемы порядка слов во временных придаточных предложениях РР. Точнее, основное внимание направлено на выяснение линейной позиции подчинительного союза. В кодифицированном литературном языке (далее: КЛЯ) союзы могут выступать только в начале придаточного предложения:

(1) ***Когда я приеду***, я тебе позвоню.¹

В РР данное правило довольно часто нарушается; например, союз может занимать позицию в середине (2) или в конце (3) придаточного:

(2) *Комаровы сидели **когда** у прокурора, они допустили целый ряд ляпсусов...* [Лаптева 1976: 316]

(3) *Я **приеду когда**, я тебе позвоню.* [ibid.: 316]

Описывая такие конструкции, О.А. Лаптева приводит примеры с двумя союзами, *когда* и *если*, см. (2)-(3) и (4):

(4) *Я буду **если** на улице Горького, я посмотрю тебе шерсть.* [ibid.: 318]

В работе О.А. Лаптевой конструкции, где союз находится в конце или в середине придаточного, трактуются как «переходные между союзным и бессоюзным подчинением». Примером конструкций с бессоюзным подчинением является следующее предложение:

(5) *Приедешь, сразу позвони.*

Бессоюзные конструкции такого типа действительно обнаруживают значение, сходное с союзами *когда* и *если*, т.е. обозначают временное следование или условие. Однако в стороне остается вопрос, возможны ли такие конструкции с другими подчинительными союзами. Если это так, то гипотеза о связи конструкций (2)-(3) с бессоюзными предложениями, аналогичными (5), не получит подтверждения.

Итак, подчинительные союзы могут занимать позицию в начале придаточного, как в КЛЯ (1), в середине (2) или в конце придаточного (3).

При анализе данного явления возникают следующие вопросы:

- какие подчинительные союзы демонстрируют такую мобильность;
- развиваются ли такие конструкции из бессоюзных предложений, аналогичных (5);

¹ Здесь и далее в примерах подчеркнуто придаточное предложение, жирным шрифтом выделен подчинительный союз.

- какова частотность в РР конструкций с подчинительным союзом в середине и в конце придаточного;
- какие факторы влияют на позицию союза (в начале / в конце / в середине придаточного).

В докладе делается попытка ответить на поставленные вопросы. Материалом исследования послужил подкорпус устной речи в Национальном корпусе русского языка (www.ruscorpora.ru), результаты Интернет-поиска на сайтах www.yandex.ru, www.google.ru, а также неопубликованные тексты устной спонтанной речи, собранные студентами МГУ и МГППУ.

В работе рассматриваются следующие подчинительные союзы: *когда, чтобы, перед тем как, после того как, пока и если*. Следует оговориться, что мы не будем рассматривать конструкции, в которых придаточное предложение находится в присловной позиции (см. [Русская грамматика 1980]), например: *теперь, когда я стал начальником*, в том числе в придаточной позиции: *...глаза... умные и твердые, когда нужно* (см. [Прокопович 1974: 114]). Кроме того, мы не рассматриваем составные союзы, частью которых являются союзы *когда, пока, чтобы и если* (*всякий раз когда; до тех пор пока; в случае если* и т.п.).

2. Частотность конструкций

Для ответа на вопросы, поставленные в п.1, был произведен подсчет частотности различных конструкций в небольшом корпусе предложений с подчинительными союзами (ок. 2 тыс. предложений). В корпус вошла выдача устного подкорпуса Национального корпуса русского языка на запрос по союзам *когда, пока, перед тем как, после того как, чтобы и если*; примеры из Интернета, полученные при помощи поисковых систем www.google и www.yandex.ru, а также неопубликованные материалы студентов МГУ и МГППУ. В результате анализа данного корпуса была выяснена приблизительная частотность различных конструкций в РР.

В среднем, наиболее частотны случаи, когда союз занимает позицию в начале придаточного предложения, т.е. конструкция, соответствующая нормам КЛЯ:

(6) ***Когда** на поинт ходили / я тоже одела каблук / а там все лужи / они замерзли / это был сплошной лёд. [НКРЯ]*

(7) *Ладно / как дела / рассказывай / **пока** время есть / а то уже скоро отправление. [НКРЯ]*

Такие примеры составляют около 85% выборки. Конструкции, где союз находится в конце придаточного, встречаются наиболее редко; таких примеров было обнаружено около 5% от общего числа. При этом распределение конструкций достаточно сильно варьирует в зависимости от конкретного союза. Наиболее часто в середине придаточного предложения выступает союз *пока*:

(8) *Он **пока** всех не обойдет не успокоится [www.chat30.ru/lenta-2008-03-30/Ангелина/]*

Такие примеры составляют более 25% из всех предложений с данным союзом. Союзы *когда* и *если* чаще всего выступают в начале придаточного, и лишь приблизительно в 15% случаев в середине придаточного:

(9) *Так он **когда** звонил / извинился и спросил / типа / разве я похож на маньяка? [НКРЯ]*

(10) *Ладно тогда / давай / **ты если** пойдёшь / напиши мне эсмэску тогда. [НКРЯ]*

Союзы *перед тем как* и *после того как* более чем в 95% случаев занимают позицию в начале придаточного предложения. Примеры на позицию в середине достаточно редки; что касается позиции в конце придаточного, такие примеры фактически единичны.

Позиция в конце придаточного, в принципе, является сильно маркированной; такие примеры составили меньше 5% выборки для *когда, если и пока* и около 2% для *после того как* и *перед тем как*.

Союз *чтобы* по распределению конструкций приближается к составным союзам: более чем в 90% случаев носители выбирают конструкцию, где *чтобы* находится в начале придаточного предложения.

3. Обобщение

Как представляется, распределение конструкций объясняется следующим. Позиция в начале придаточного предложения является для подчинительных союзов немаркированной, данные примеры наиболее частотны. Настоящий принцип верен не только относительно материала русского языка, но и подтверждается на материале различных генетически неродственных языков (см. Hawkins 1983). Эту закономерность принято объяснять факторами когнитивного и дискурсивного плана: во-первых, более «тяжелая» составляющая (т.е. придаточное) имеет тенденцию занимать позицию справа от более «легкой»; во-вторых, говорящий стремится сразу же подать слушающему сигнал о том, как интерпретировать последующую информацию, точнее, является ли она тем, что непосредственно утверждается (главное предложение) или подается как фоновая (придаточное). Для этого необходимо в первую очередь вербализовать подчинительный союз.

Тем не менее, в РР это правило нарушается. В каких случаях оно может нарушаться? В результате анализа примеров из полученного нами корпуса, можно сделать следующие выводы. В середине или в конце придаточного союз чаще всего выступает в тех случаях, когда само придаточное находится в препозиции к главному. В результате распределение конструкций различно у союзов *когда*, *пока*, *если* vs. *чтобы*. Действительно, *когда*, *пока* и *если* чаще всего вводят придаточное в препозиции, в то время как целевые придаточные с союзом *чтобы* чаще всего занимают позицию после главного предложения (см. [Русская грамматика 1980]).

Как представляется, в данном случае оказывает влияние характерное для РР правило выноса важной информации в начало предложения, которое было отмечено в [Земская 1987] и др. работах. Союз занимает позицию в середине или в конце придаточного, если для говорящего высоко релевантными оказываются какие-либо участники ситуации или сам предикат. Например, рассмотрим следующие примеры:

(11) *Вот сейчас посмотрите / даже Радуева после того как поймали / какой милый стал.* [НКРЯ]

В данном примере существительное *Радуева* с частицей *даже* находится в фокусе ремы. Однако неверное было бы предположить, что вынос данного элемента перед союзом определяется его рематичностью. Дело в том, что выступать в данной позиции может также топик, или часть темы, например:

(12) *Ну / вот у меня гастрит чтобы проверить / посадили на диету.* [НКРЯ]

В (12) *гастрит* входит в состав темы, в то время как *у меня* является контрастивным топиком. При этом обе группы выносятся в позицию перед союзом.

Можно заключить, что конструкции, рассмотренные в (11)-(12) – проявление более общего для РР явления выноса в начало предложения информации, релевантной для говорящего. При этом имеется в виду именно дискурсивная релевантность, а не принадлежность теме или реме высказывания. Как представляется, аналогичное явление имеет место в конструкциях с разрывом именной группы, например:

(13) *Мхом пахнет прелым и грибами.*

(14) *В пакетах дайте мне, пожалуйста, молоко.* [Крылова, Хавронина 1984: 138]

Подобные конструкции РР обсуждаются в ряде исследований (см. [Лаптева 1976], [Земская 1987] и др.). Ср. тж. [Крылова, Хавронина 1984], где подобные примеры трактуются в терминах нарушения порядка слов в «эмоционально окрашенной речи».

Отдельно следует рассматривать примеры, в которых союз занимает позицию после местоимения (обычно безударного), ср. (8)-(10) (такие примеры выделяются особо также в работе [Лаптева 1976: 315]). Союз вместе с местоимением составляют в таких предложениях одно фонетическое слово. Трудно сказать, что местоимение в таких конструкциях является дискурсивно выделенным; отметим, что на него не падает логическое ударение (в отличие от *у меня* в (12)). Как представляется, данные конструкции отражают тенденцию местоимений и личных показателей в различных языках занимать позицию в непосредственном начале предложения, независимо от их синтаксической позиции. (См. [Givón 1990].)

Далее, релевантным оказывается фактор «тяжести» союза, т.е. идет ли речь о простом или составном союзе. Как представляется, этот фактор обуславливает особое поведение составных союзов *после того как* и *перед тем как*: данные союзы проявляют гораздо меньшую мобильность, чем простые союзы *когда*, *если* и *пока*.

Интересно, однако, что союз *чтобы* демонстрирует те же свойства, что и составные союзы. Однако это объясняется следующим. *Чтобы* может использоваться как в целевых обстоятельственных предложениях, так и в изъяснительных предложениях при глаголах *везти*, *собираться*, *ждать* и т.д. Если рассматривать данные употребления отдельно, получается следующее распределение: в середине и в конце предложения может выступать только обстоятельственное *чтобы*:

(15) Ну как / это мы и сейчас поедим / и назавтра **чтобы** осталось. [НКРЯ]

(16) [А., жен] Там была большая насыпь / там детей клали **чтобы** / так шуму не было. [Китайгородская, Розанова 1964]

Изъяснительный союз *чтобы* во всех примерах выборки занимает позицию в начале предложения. Если не учитывать примеры с *чтобы* в изъяснительных придаточных, распределение частотности конструкций с обстоятельственным *чтобы* в большей степени соответствует результатам по остальным простым союзам.

Интересно также обратить внимание на тот факт, что среди простых союзов наиболее часто в середине или в конце придаточного предложения может выступать союз *пока*. Как объясняется особое поведение данного союза? Дело в том, что союз *пока* в СРЯ омонимичен наречию *пока* со значением «в настоящий момент, сейчас». Наречие *пока* может достаточно свободно передвигаться в пределах предикации, занимая позицию как справа от глагола, так и в конце предложения, ср.:

(17) [№ 2, муж] Я имел неосторожность закончить филфак / *учусь пока* в аспирантуре / но сейчас устроился на радиостанцию музыкальным. [НКРЯ]

(18) [№ 3, муж] Я хирург / *хорошего ничего сказать не могу про власть пока* / даже касаясь хирургии. [НКРЯ]

Примеры (17) и (18) допускают различные изменения в позиции *пока*: *пока учусь в аспирантуре*; *учусь в аспирантуре пока*; *хорошего ничего пока сказать не могу про власть* и т.д. Как представляется, синтаксические свойства наречия в какой-то мере влияют на свойства союза *пока* в РР. Отметим, что другие союзы (*когда*, *если* и *чтобы*) не имеют подобных омонимичных пар и являются менее «мобильными» в РР.

4. Заключение

Итак, в работе анализируются вопросы порядка слов в обстоятельственных предложениях устной спонтанной речи. В результате анализа данных корпуса РР можно сделать следующие выводы:

Конструкции, в которых союз занимает позицию в середине или в конце придаточного предложения, могут возникать не только с союзами *когда* и *если*, но и со следующими обстоятельственными союзами: *чтобы*, *пока*, *перед тем как*, *после того как*. Такие значения невозможны в бессоюзных предложениях, аналогичных (5). Следовательно, гипотеза О.А.

Лаптевой о связи конструкций (2)-(3) с бессоюзными предложениями не получает подтверждения.

Наиболее частотными в РР являются конструкции, где подчинительный союз занимает позицию в начале придаточного предложения. Однако в случае, если говорящий хочет особо выделить какие-либо компоненты придаточного, данные компоненты занимают позицию перед союзом. Местоимения могут выноситься в позицию перед союзом вне зависимости от их дискурсивной выделенности.

Из всех исследованных союзов чаще всего позицию в середине или в конце придаточного предложения занимает союз *пока*, что, вероятно, обусловлено наличием наречия *пока*, которое не занимает в предложении фиксированную позицию. Наиболее редко допускается вынос каких-либо компонентов перед составными союзами *перед тем как* и *после того как*.

Литература

1. Валгина Н.С. *Синтаксис современного русского языка*. М.: «Высшая школа».
2. Галкина-Федорук Е.М., Горшкова К.В., Шанский И.М. 1958. *Современный русский язык. Синтаксис*. М.: Государственное учебно-педагогическое издательство министерства просвещения РСФСР.
3. Земская Е.А. 1987. *Русская разговорная речь: лингвистический анализ и проблемы обучения*. М.: «Русский язык».
4. Китайгородская М.В., Розанова Н.Н. 1999. *Речь москвичей: Коммуникативно-культурологический аспект*. М.
5. Крылова О.А., Хавронина С.А. 1984. *Порядок слов в русском языке*. М.: «Русский язык».
6. Лаптева О.А. 1976. *Русский разговорный синтаксис*. М.: «Наука».
7. Прокопович Н.Н. 1974. *Вопросы синтаксиса русского языка*. М.: «Высшая школа».
8. Русская грамматика 1980. ИРЯ АН СССР. М.: «Наука».
9. Givón T. 1990. *Syntax: a Functional-Typological Introduction*, v. 2. Amsterdam/Philadelphia: John Benjamins.
10. Hawkins J.A. 1983. *Word Order Universals*. New York: Academic Press.

Н.М. Шитова

Относительное предложение в кодифицированном литературном языке и в разговорной речи

*МГУ им. Ломоносова,
г. Москва, Россия
natalia-shitova@yandex.ru*

Введение

Данная работа посвящена сопоставительному исследованию типов относительных предложений в кодифицированном литературном языке (далее КЛЯ) и в русской разговорной речи (далее РРР). Вначале приводится попытка исчисления возможных типов относительных предложений в КЛЯ и РРР, далее рассматривается их сравнительная частотность в корпусе устных текстов (монологических и диалогических).

В исчислении типов относительных предложений используются следующие обозначения: [...] — главное предложение, (...) — придаточное предложение, NP, D — соответствующие составляющие. Для каждого типа приводится пример из корпуса текстов.

1. Исчисление типов относительных предложений в КЛЯ

Тип 1

Данный тип является наиболее частотным в КЛЯ (как, впрочем, и в РРР, что подтверждается статистикой, приводимой ниже). В КЛЯ можно выделить следующие подтипы:

1) [NP_i...](который_i...)

(1) *Комната, в которую вступил Иван Иванович, была совершенно пуста (Гоголь)¹*

2) [D_i...](что/кто_i...), причем в D нет N

(2) *Тот, кто играет словами, не всегда играет чувствами (Лерм.)²*

Заметим в связи с первыми двумя подтипами, что логически возможные конструкции [NP_i...](что/кто_i...) и [D_i...](который_i...) в литературном языке нормативными не считаются (об этом ниже при рассмотрении типов относительных предложений в разговорной речи).

3) [NP_i...](где/когда_i...)

(3) *В городе, где он жил, имелись два кинотеатра плюс четыре дома культуры. [Сахновский Игорь. Ревнивый бог случайностей // "Октябрь", 2003]*

4) [D_i...](где/когда_i...), причем в D нет N

(4) *Там, где люди вместе работают на одну цель, можно решить любую проблему. [Александр Градов. "Медведи" залили партийный фундамент // "Аргументы и факты", 2003.01.22]*

5) [NP...](какой...)

(5) *По вечерам появлялись на улицах страшные люди, каких раньше никто и не видывал (А. Н. Толст.)³*

Легко видеть, что семантика местоимения *какой* существенно отличается от семантики местоимений и местоименных наречий, используемых в предыдущих подтипах. Как отмечается в РГ-80, «как средство оформления анафорической связи слово *какой* употребляется в

¹ Пример из [РГ-80]

² Пример из [РГ-80]

³ Пример из [РГ-80]

тех случаях, когда лицо или предмет, названные существительным, характеризуются через их отношение к другим лицам или предметам того же класса (или множества), причем само это отношение является отношением соответствия, подобия или сходства». Заметим, что данное отличие не мешает нам рассматривать данную конструкцию в рамках Типа 1, т.к. она обладает признаками, объединяющими данный тип предложений и отличающими его от других типов, - препозиция главной части по отношению к придаточной и редуцированного способа выражения кореферентной именной группы в придаточном (ср. тип 2). Таким образом, подтипы Типа 1 объединяются на основе формальных признаков и имеют различия в семантическом аспекте: подтипы 1) и 2) отличает «предметная» ориентация, подтипы 3) и 4) — «пространственно-временная», подтип 5) — «признаковая».

Тип 2

[NP_{1i...}](каковой NP_{2i...})

(6)...у буфетчика, уже знавшего из газет о гибели Берлиоза и о месте его проживания, мелькнула мысль о том, что уж не служили ли, чего доброго, по Берлиозу церковную панихиду, каковую мысль, впрочем, он тут же отогнал от себя, как заведомо нелепую. [М.Булгаков. Мастер и Маргарита, часть 1 (1929-1940)]

Нельзя не отметить редкую встречаемость данного типа. По нашим численным данным, конструкция встретилась в НКРЯ всего 202 раза. Интересно заметить, что данный тип не является однородным в точки зрения употребления. Можно выделить 2 подтипа:

1) NP₁ и NP₂ совпадают. (см. пример (6))

2) NP₁ и NP₂ не совпадают, однако NP₂ является квазисинонимом с более общим значением:

(7)Сейчас можно говорить скорее о перспективе ускоренного проникновения относительно респектабельных форм православного национализма в "большую политику", для какового процесса радикалы будут выполнять лишь необходимую функцию маргинальной группы поддержки. [Александр Верховский. Православный национализм. Заметки об определении и о перспективах движения // "Неприкосновенный запас", 2003]

Интерес представляет также ряд примеров особого вида (объединим их условно в подтип 3) Типа 2), в которых NP₂ соотносится не с отдельно взятой NP₁ в составе главного предложения, а со всем главным предложением.

(8)В это же самое время, как говорилось, Лютославский мучительно ищет новые методы упорядочения 12-тоновой вертикали, каковые поиски к концу 50-х сложатся в его уникальный стиль. [Витольд Лютославский: классик XX столетия // "Российская музыкальная газета", 2003.02.12]

Нетрудно видеть, что такие употребления являются аналогами таких предложений, которые можно опять же условно приписать Типу 1.

(9)Он опоздал, что было замечено.⁴

Следуя А. А. Зализняку и Е. В. Падучевой, не будем подробно останавливаться на такого рода предложениях, выделенных в статье [Зализняк, Падучева 1975] в «особый тип придаточных относительных». Отметим лишь, что для КЛЯ данный подтип не является редким. НКРЯ позволяет продемонстрировать следующую статистику для Типа 2 в КЛЯ.

Подтип	Число примеров в данном подтипе	Процент примеров данного подтипа в КЛЯ
1) ⁵	69	34%
2)	73	36%
3)	60	30%

⁴ Пример из [Зализняк, Падучева 1975]

⁵ В данной статистике учитывались также примеры парцелляции. Например, «Арабский язык железно замкнут в своем регионе. Каковой регион стремительно деградирует, а вовсе не развивается».

2. Исчисление типов относительных предложений в РРР.

Тип 1 (соответствует Типу 1 в КЛЯ)

Данный тип относительных предложений является наиболее распространенным в РРР.

- 1) [NP_i...](который_i...)
 (10) Надо, вот, просто браться только за те дела, которые мне по силам.
- 2) [D_i...](что/кто_i...), причем в D нет N
 (11) Так, только то, что я на доске прикрепляю, мы на это не смотрим.
- 3) [NP_i...](что/кто_i...)
 (12) Дойди до киоска, что стоит на углу⁶.
- 4) [D_i...](который_i...), причем в D нет N
 (13) Так вот, а куда деваться тем, которые не гераклы?

Можно заметить, что, по сравнению с литературным языком, разговорная речь демонстрирует все логически возможные употребления анафорических местоимений в придаточном. В связи с рассмотрением данного типа относительных придаточных нельзя не отметить проблему дистрибуции «который» и «кто, что» в РРР.

Как отмечается в статье [Янович, Грунтова 2007], «употребления *кто (что)* и *который* дополнительно распределены по контекстам в зависимости от наличия лексического имени в вершине (в КЛЯ — Н. Ш.). *Кто, что* оформляют относительное предложение, если его вершина пуста либо в ней есть местоимения типа *любой, кто-то, все* и под. (D), и нет имени (N). Употребление местоимения *который* возможно только при наличии имени (N) в вершине (DP)». Таким образом, конструкции 3) и 4) «незаконны» с точки зрения КЛЯ. Однако, как можно будет видеть ниже, РРР допускает такие употребления в достаточно большом проценте случаев (в сумме для 3) и 4) 17,03%⁷). Авторы рассматриваемой статьи указывают на тот факт, что большинство примеров конструкции «D + *который*» относится к текстам XVIII - XIX вв.

- (14) *Так и все, которые в делании добродетели ленижны, суть младенцы.* [Платон (Левшин), архиепископ Московский и Калужский. (1779)]⁸

В РГ-80 отмечается «возможность употребления на месте слова *который* местоимения-существительного *кто* (при отсылке к лицу) и *что* (при отсылке к предмету)».

- (15) *Плюшкин сел в кресла и уже, казалось, не мог найти материю, о чем говорить* (Гоголь)⁹.

Однако примеры, иллюстрирующие данное явление, относятся преимущественно к литературе XIX-н.ХХ веков. Таким образом, можно утверждать, что в РРР встречаются конструкции, свойственные скорее языку XIX в., чем современному КЛЯ.

Вернемся к исчислению.

- 5) [NP_i...](где/когда_i...)
 (16) Представляешь, это единственное место в Паттайе, где на закуску дают соленые огурцы.
- 6) [D_i...](где/когда_i...), причем в D нет N
 (17) Мне нравится *там / где* я живу. [Беседа в Новосибирске (2003.07.22)]

⁶ Пример из [Янович, Грунтова 2007]

⁷ К сожалению, далеко не все примеры однозначны с точки зрения отсутствия в них эллипсиса (имеются в виду случаи типа «Он мою подругу оскорбил. Она самая добрая из тех, что я встречал»). Там не менее, процент таких оборотов в РРР существенно выше соответствующих данных для НКРЯ (в сумме 194 примера на корпус, считая случаи явного эллипсиса [данные из Янович, Грунтова 2007]).

⁸ Пример из [Янович, Грунтова 2007].

⁹ Пример из РГ-80.

Тип 2 (тип 1в РРР с «инверсией подчинения»¹⁰)

- 1) (который_i...) [NP_i...]
(18) *Которому* вышака, тот там и остаётся, а *если* вывели, то два го́да. [Юрий Домбровский. Факультет ненужных вещей, часть 4 (1978)]¹¹
- 2) (что/кто_i...) [D_i...], причем в D нет N
(19) Кто работает за свои, за эти самые...кровные, тот ничего.
- 3) (что/кто_i...) [NP_i...]
(20) У кого поджелудочная железа и печень больные, те пьют панкреатин.¹²
- 4) (который_i...) [D_i...], причем в D нет N
(21) *Которые тут самые умные, те пусть сами разбираются.*

Заметим, что данный тип мало распространен в литературном языке, за исключением подтипа 2), который, впрочем, также не может считаться общеупотребительным. Абсолютное большинство употреблений в художественной литературе, приводимых в НКРЯ, являются примерами стилизации.

Тип 3 (Тип 1 в РРР с эллипсисом NP/D в главной части)

- 1) [∅_i...](который_i...)
(22) *Поднимите руку которые в понедельник сдавать хотят.*
- 2) [∅_i...](что/кто_i...)
(23) *За проезд прошу платить в кассу кто еще не платил.*¹³

Тип 4 (Тип 2 в РРР с эллипсисом NP/D в главной части)

- 1) (который_i...) [∅_i...]
(24) *Которые на площади вошли оплачивайте.*
- 2) (что/кто_i...) [∅_i...]
(25) *Кто поопытнее захватывал места поудобнее. (А.Лаптев, В пути)*¹⁴

Типы 3 и 4 относительных предложений подробно описаны О. А. Лаптевой (см. [Лаптева 1976, 352-362]).

Рассмотренные типы относительных предложений характеризуются разной частотой встречаемости в РРР. Ниже приводятся результаты исследования на основе небольшого (375 относительных предложений) корпуса устных текстов, представленного такими жанрами, как бытовой диалог и лекция. Итоговая таблица отражает частотность выделенных типов относительных предложений, а также частотность причастных оборотов в разговорной речи.

¹⁰ Термин из [Зализняк, Падучева 1075]

¹¹ Демонстрируя данный подтип, мы считаем возможным привести пример из художественной литературы как случай стилизации

¹² Пример из [Лаптева 1976]

¹³ Пример из [Лаптева 1976]

¹⁴ Пример из [Лаптева 1976]

3. Статистика по относительным оборотам в РРР

Таблица 1.

ТИП	ПОДТИП	КОНСТРУКЦИЯ	ПРОЦЕНТ В ПОДКОРПУСЕ (375 ед.)
1	1)	[NP _i ...](который _i ...)	63,3%
	2)	[D _i ...](что/кто _i ...)	13,3%
	3)	[NP _i ...](что/кто _i ...)	1,8%
	4)	[D _i ...](который _i ...)	4,23%
	5)	[NP _i ...](где/когда _i ...)	4,5%
	6)	[D _i ...](что/кто _i ...)	0
2	1)	(который _i ...) [NP _i ...]	0
	2)	(что/кто _i ...) [D _i ...]	1,33%
	3)	(что/кто _i ...) [NP _i ...]	0
	4)	(который _i ...) [D _i ...]	0
3	1)	[ø _i ...](который _i ...)	0,27%
	2)	[ø _i ...](что/кто _i ...)	0
4	1)	(который _i ...) [ø _i ...]	0
	2)	(что/кто _i ...) [ø _i ...]	0,27%
Причастный об.			12%

Таким образом, наиболее частотной является общая для РРР и КЛЯ конструкция [NP_i...](который_i...). К сожалению, мы не обладаем аналогичными статистическими данными для литературного языка, однако осмелимся предположить, что процент употребления данной конструкции в КЛЯ будет примерно равен соответствующей цифре для разговорной речи. Это связано, на наш взгляд, с одновременным действием двух факторов. Так, с одной стороны, литературный язык демонстрирует меньшее разнообразие типов относительных предложений, по сравнению с разговорной речью, но с другой стороны, процент причастных оборотов в КЛЯ существенно выше, чем в РРР.

Прочие конструкции являются значительно менее распространенными.

4. О различиях относительных оборотов в КЛЯ и РРР

О расположении частей относительных предложений.

КЛЯ не допускает относительных конструкций, в которых придаточная часть находится в препозиции к главной, «в художественной литературе подобные конструкции отмечаются только как прием стилизации»¹⁵. В то же время, в РРР такие конструкции встречаются, хотя и не могут быть признаны широко употребительными. А. А. Зализняк и Е. В. Падучева отмечают, что данный вид относительных предложений (модель Б2 в их классификации, Типы 2 и 4 в представленном исчислении) является основным в частности для древнерусского и существует в настоящее время «в русской диалектной речи и в просторечии» [Зализняк, Падучева 1975, 69].

О месте местоимения который в относительном придаточном

В литературном языке относительное местоимение *который* располагаются в начале придаточного предложения [РГ-80, § 2881], за исключением случаев так называемого Эф-

¹⁵ Авторы РГ-80 отмечают необходимость препозиции придаточного в случае «введения в его состав существительного, с которым синтаксически связано местоименное слово и относительно которого это слово располагается также препозитивно: *На котором месте школа стоит, пустырь был* (Бажов)». Т.к. такого рода придаточные не являются относительными в строгом смысле, а принадлежат к типу коррелятивных, то отдельному рассмотрению в данной работе они не подвергаются.

фекта Крысолова, когда на левую периферию придаточного смещается не само местоимение, а объемлющая его составляющая:

Как отмечается в РГ-80 и в [Тестелец 2001], конструкции, нарушающие данное правило в современном КЛЯ не употребляются, хотя признавались грамматичными в XIX веке.

(26) *Неужели я принадлежу к числу тех людей, которых один вид уже порождает недоброжелательство (Лермонтов)*¹⁶

В РРР также возможно несоблюдение данного правила:

(27) *Там книги лежали, которых...ну названия мне ничего не говорят.*¹⁷

К отличие от литературного языка, РРР допускает употребление относительного местоимения *который* в конце или середине придаточного, что связано с топикализацией некоторого другого члена.

(28) *Купила билет на сеанс, всего лишь через тридцать минут который шел.*

Таким образом, возможность употребления подтипов 3) и 4) Типа 1 в РРР (конструкции [NP_i...](что/кто_i...) и [D_i...](который_i...)), а также допустимость так называемой инверсии подчинения (Типы 2 и 4) и более свободного, нежели в КЛЯ, расположения подчинительного местоимения в придаточном, подтверждают мысль об общей близости РРР к диалектному, сохранившему древние особенности языку.

Литература

1. Edward L. Keenan. Relative clauses. Language typology and syntactic description. Vol 2. 1985.
2. Зализняк А. А., Падучева Е. В. К типологии относительного предложения. Семиотика и информатика: Сб. научн. статей/РАН. М., ВИНТИ, 1975.
3. Русская грамматика. М., 1982.
4. Лаптева О.А. Русский разговорный синтаксис. М., Наука, 1976.
5. Тестелец Я.Г. Введение в общий синтаксис. М., 2001.
6. Янович И., Грунтова Л. Дистрибуция русских относительных местоимений *КТО(ЧТО...)* vs. *КОТОРЫЙ*. Труды международной конференции “Диалог 2007”.

¹⁶ Пример из РГ-80

¹⁷ Данный пример не заслуживает абсолютного доверия, т.к. демонстрирует явление коррекции.

М.С. Шматова

Образование отглагольных существительных на -ка от глаголов несовершенного вида в русской устной речи

*МГУ им. Ломоносова,
г. Москва, Россия
mashashma@gmail.com*

Введение

Хорошо известно, что устная спонтанная речь характеризуется рядом особенностей, отсутствующих в кодифицированном литературном языке (далее КЛЯ). Эти различия зафиксированы на разных уровнях языка. Например, нарушение порядка слов и согласования, повторы, частое употребление местоимений, наличие большого количества значений у одной лексемы и т.п. Эти и другие особенности описаны в работах Е.А.Земской, О.А. Лаптевой, М.В. Китайгородской и Н.Н. Розановой.

Данная статья посвящена исследованию образования существительных, оканчивающихся на -ка, от различных глаголов КЛЯ в устной речи. Ниже будут рассмотрены следующие вопросы:

1. от каких семантических классов глаголов несовершенного вида возможно образование существительных на -ка в устной речи
2. от каких глагольных основ и в каких случаях возможно образование рассматриваемых существительных;
3. как семантически различаются отглагольные существительные на -ка по отношению к глаголам, от которых они образованы;
4. насколько регулярно образование таких существительных в рамках того или иного семантического класса предикатов.

В исследовании были использованы материалы подкорпуса устной речи национального корпуса русского языка, диктофонные записи диалогов, личные наблюдения, а также материалы форумов, он-лайн блогов и icq-клиентов.

1. Общие замечания

Для исследования была взята глаголы из 12 семантических классов по фундаментальной классификации предикатов Ю.Д. Апресяна: действия, деятельности, занятия, воздействия, проявления, положения, состояния, свойства, параметры, существования, оценки и интерпретации.

Внутри каждого класса рассматривалось от одного до трех типичных глаголов. Стоит сразу оговориться, что не были рассмотрены глаголы на –ся, которые могут употребляться и без –ся, так как в таком случае не является очевидным, от какого из двух глаголов (на –ся или на –ть) образовано существительное (так, в классе «свойства» не был рассмотрен типичный глагол *виться* (ср. *вить*), тогда как в классе «свойства» присутствует глагол *заикаться* (ср. **заикать*)).

В ходе исследования рассматривались только существительные, образованные от глаголов несовершенного вида. Стоит отметить, что в некоторых случаях у рассматриваемых глаголов существуют видовая пара, и, очевидно, в таких случаях отглагольные существительные со значением результата стоит интерпретировать как образованные от глагола совершенного вида (например, *ошибка* от *ошибиться*, а не от *ошибаться*), а таковые в данном исследовании не рассматриваются.

2. Особенности образования отглагольных существительных на -ка

Можно выделить несколько стратегий образования исследуемых существительных в русском языке.

Существительные на -ка могут образовываться от:

основы инфинитива (по классификации А.В. Исаченко [2]) при помощи следующих стратегий:

1. Присоединение -ка к основе инфинитива: писа-ка
2. Присоединение –л-, а затем -ка к основе инфинитива: писа-л-ка
3. Присоединение –н-, а затем -ка к основе инфинитива: лежа-н-ка
4. основы настоящего времени с помощью присоединения -ка: стой-ка

В ходе исследований было замечено, что продуктивной для образования новых существительных в устной речи является вторая из рассмотренных стратегий (а именно «инф + -лка»). Все полученные в ходе работы существительные, образованные по другим моделям, встречаются в КЛЯ. Поэтому предметом рассмотрения данной статьи являются собственно окказионализмы на -лка.

В Русской Грамматике 1980 г. –лк(а) выделяется в качестве отдельного словообразовательного суффикса. Также отмечается, что существительные на –лк(а) «называют предмет (преимущественно неодушевленный), производящий действие, названное мотивирующим словом, или предназначенный для выполнения этого действия». Как мы увидим далее, в устной речи такие существительные появляются и в других значениях.

Были выяснены следующие ограничения на образование отглагольных существительных на -лка:

1. глаголы с суффиксом -ну не могут образовывать отглагольные существительные на -ка.
2. глаголы, основа инфинитива которых состоит более чем из трех слогов, чаще всего не образуют существительные на -ка. При просмотре корпуса и блогов был найден только один пример на подобное образование: существовать > существовалка.

3. Стилистические особенности

Наиболее продуктивна рассматриваемая модель образования в области компьютерной и технической разговорной лексики. Скорее всего, это происходит, так как большая часть наименований компьютерных деталей — заимствованные из английского слова, которые слишком неудобно употреблять в русской обиходной речи или же длинные и неудобные в употреблении. Именно поэтому происходит название того или иного «гаджета» или компьютерной игры по его/ее основной функции. Например, писалка вместо CD-Writer или записывающее устройство, ходилка вместо action-adventure.

4. Семантические особенности

Всего было рассмотрено 23 глагола, для каждого из которых в результате поиска по разным источникам (см. выше) было получено по одному отглагольному существительному, образованному по модели «инф+-лка». При этом, многие окказионализмы имеют несколько значений.

Приведем здесь сравнительную таблицу возможных значений окказионализмов:

Таблица 1. Возможные значения окказионализмов

Значение	Количество	Процентная доля
Субъект	16	42,1%
Объект	1	2,6%
Инструмент	10	26,3%
Место	2	5,3%
Ситуация	3	7,9%
Результат	1	2,6%
Лесикализация (в компьютерной лексике)	5	13,2%

Видно, что кроме выделенных для КЛЯ в Русской Грамматике значений для существительных на –лка (субъект и инструмент), присутствуют и другие, характерные собственно для устной речи.

Теперь рассмотрим существительные с каждым из полученных значений.

4.1. Существительные со значением субъекта

Как видно из таблицы, отглагольных существительных на -ка со значением субъекта действия получено больше всего. Они могут образовываться от глаголов всех рассмотренных классов. Возможность образования от конкретного глагола, видимо, полностью зависит от его индивидуальных семантических свойств и употребляемости.

Такие существительные могут обозначать:

1) одушевленный субъект

а) не важен пол субъекта

А делалка-то у вас только одна — я.

(2) Подводилка ты! ВОТ!))))))))

Я? ;) Это вы меня подвели... \

б) субъект только женского пола

(3) Курортные гулялки (т.е. женщины, приезжающие на курорт гулять и знакомиться с мужчинами) обитают чаще всего парами. [женский форум]

2) неодушевленный субъект

(4) Твоя сумка — чудесная вместила всего на свете. [разговор]

Интересно также, что всякие новообразования на -ка возникают в качестве различных никнеймов на форумах и в блогах. Были найдены такие «сетевые имена», как Светилка, Хотелка, Бывалка. Видимо, причина тут в желании обладателя ника показаться оригинальным («такого больше ни у кого нет, потому что такого слова в словаре нет»), а не в какой-то тонкой новой семантике данного образования.

У некоторых таких существительных также существует эквивалент в КЛЯ:

(5) старый системник-десктоп с наклейкой ORD, стоялка с компакт-дисками, подставка под чашку из пимпочек от пробок 20л бутылей с водой и компакт-дисков. [LiveJournal]

В данном примере стоялка имеет эквивалент стойка. Возможно, такая замена происходит, потому что модель образования «инф+–лка» является наиболее частотной для образования новых отглагольных существительных в устной речи, а значит, наиболее удобная, и поэтому имевшиеся значения тоже «пересматриваются» с точки зрения удобства.

4.2. Существительные со значением объекта

Похоже, что данное значение у таких существительных отсутствует. Встретился найден только один пример существительного с таким значением, но и он сомнителен:

PS а ещё здесь есть мои хотелки по поводу путешествий в другие города.;). [community.livejournal.com/hotelka]

Ну и конечно же побольше разных радующих душу моточков пряжи и вдохновения для воплощения вязальных хотелок-мечталок [forum.hobbyportal.ru]

Здесь хотелка трактуется как «что-то, которое хочет индивид». Не случайно в трактовке употреблено «что-то». По найденным контекстам можно интерпретировать это «что-то» как объект, а можно как абстрактное понятие (так, в (7) для словосочетания вязальные хотелки возможна интерпретация «желание связать что-либо», а возможна - «вязанный объект, который хочет создать говорящий»).

Пример может являться также иллюстрацией встречающегося в устной речи употребления отглагольных существительных, образованных по модели «инф+–лка» в тех же значениях, что и отглагольные существительные на –ние в РЛЯ (в (6) хотелка вместо хотение).

4.3. Существительные со значением инструмента

Это вторая по величине группа рассмотренных существительных. Такие существительные могут образовываться от глаголов всех семантических классов, кроме классов воздействия, существования и положения. Здесь также можно выделить два подкласса:

1) Физический объект, с помощью которого совершается то или иное действие:

(8) Писалка (т.е. пишущий CD-ROM) перестала видеть диски и писать.

2) Название способности:

(9) понималка (т.е. способность понимать) развивается немислимыми темпами, када есть чего бояться в сфере чувств.....када есть чего бояться, заглядывая в зеркало своей души....в .. [forum.psy.net.ru]

4.4. Существительные со значением места

Было замечено, что окказионализмы с данным значением могут образовываться только от одного семантического класса глаголов — положения. Также данное значение можно определить не просто как значение места, а как значение места, совмещенное со значением инструмента. Рассмотрим это на конкретных полученных примерах:

(10) в коляске поднимается лежалка, и я стала ее иногда поднимать, когда лежать устает.

Здесь под «лежалкой» первоначально понимается приспособление, с помощью которого «можно лежать», то есть, которое позволяет принять удобное положение лёжа. В данном случае это часть детской прогулочной коляски.

(11) В поход с собой необходимо взять еще дождевик и плотную сиделку. (инструктаж в турклубе)

В данном примере сиделка — это объект, на котором (или же при помощи которого, т.е. интерпретация здесь неоднозначная) можно находиться в положении сидя в походных условиях (на мокрой траве, к примеру).

Как видно из анализа приведенных примеров, это значение места является производным от значения инструмента, и сохраняет в себе черты второго.

4.5. Существительные со значением ситуации

Некоторые существительные этой группы используются вместо принятых в КЛЯ отглагольных существительных на –ние.

(12) но на системе 85W 16 — 25000 Гц с хорошо подогнанным эквалайзером звучалка (т.е. звучание) будет лучше

Глаголы, от которых в КЛЯ нет образований на –ние со значением состояния, в устной речи приобретают их посредством той же стратегии.

(13) ... паворот и плавно таки вернутся в левый ряд, имхо мне это проще чем играть в уступалки (= *уступание), темболее что обгон слева запрещен. [tourist.kharkov.ua/phpbb]

4.6. Существительные со значением результата

На такое значение был обнаружен только один пример, так как семантически оно скорее является характерным для существительных, образованных от глаголов совершенного вида.

(14) По-крайней мере первую игру хочу посмореть штоб понять ошибалки свои [bloodyhawks.ru/gb]

Здесь ошибалка является эквивалентом ошибки.

Скорее всего, данное значение только начинает появляться у существительных, образованных от глаголов несовершенного вида в силу продуктивности в устной речи рассматриваемой модели.

4.7. Лексикализация в сфере компьютерной терминологии

Как уже было сказано выше, область компьютерной лексики является продуктивной для образования отглагольных существительных на -ка. Многие образованные окказионализмы уже лексикализировались. В особенности это касается «народных» названий компьютерных игр, которые постепенно закрепляются в разговорной речи.

(15) 1: У меня появилась ходилка новая!

2: Какая это?

[разговор в icq]

(16) 1: Чё делаешь?

2: Гамаю в последнюю торговалку (вид экономической игры)

[разговор в icq]

4.8. Многозначные отглагольные существительные

Многие рассмотренные отглагольные существительные на -ка, как уже было упомянуто выше, имеют более, чем одно значение. Ниже приведем сопоставительную таблицу таких значений с примерами.

Таблица 2. Значения многозначных существительных на -ка

Существительное	Значения	Примеры
ходилка	субъект	На вокзале игрушки продают... ну, звери... ходилки такие
	лексикализация	см. (14)
делалка	субъект	см. (1)
	инструмент	Да у нее же эта, как ее... делалка коктейлей есть
торговалка	инструмент	Хорошая торговалка (если относиться к ней как к торговому терминалу). Меньше всех трафик жрет. (= программа для торгов на бирже)
	лексикализация	см. (16)
звучалка	субъект	Твоя звучалка на весь дом орет, просыпайся!
	инструмент	Никак не смог установить звучалку. Ком даже ее не видит. (звуковая карта)
	лексикализация	Звучалка по-прежнему хромает. (часть программы, отвечающая за звук)
	ситуация	см. (12)
хотелка	инструмент	а моя хотелка хочет 350D (название модели фотоаппарата)
	объект	см. (6), (7)
уступалка	субъект	вот вы уступалки))))мне пока не надо..... у меня и так хватает что доделать и переделать ;)))) [www.coupeclub.ru/forum]
	ситуация	см. (13)
ошибалка	субъект	...потому что вы, уважаемые мои, все равно часто ошибаетесь, оставляя прогноз, ошибалки вы эдакие. ... [LiveJournal]
	результат	см. (14)

5. Заключение

В данной статье были рассмотрены основные способы образования и значения отглагольных существительных на -лка в устной речи. На основании вышесказанного можно сделать несколько выводов:

1) В последнее время образование отглагольных существительных является достаточно продуктивной моделью.

2) Отглагольное существительное может иметь больше значений, чем подобные образования в КЛЯ.

3) Существует также зависимость образования разных значений отглагольных существительных от семантического класса глагола. Так, существительные со значением «место+инструмент» могут образовываться только от глаголов класса положения, а существительные со значением субъекта, в свою очередь, от таких глаголов не образуются, но образуются от глаголов всех остальных рассмотренных классов. Лексикализация происходит только от глаголов классов действия, деятельности и проявления.

Литература

1. Апресян Ю.Д. Фундаментальная классификация предикатов и системная лексикография. Грамматические категории: иерархии, связи, взаимодействие. СПб.: 2006
2. Винокур Т.Г. Стилистическое развитие современной русской разговорной речи. – В кн.: Развитие функциональных стилей современного русского языка. М., 1968 Русская грамматика т.1, М., Наука, 1980
3. Земская Е.А. Русская разговорная речь: лингвистический анализ и проблемы обучения. М., «Русский язык», 1987
4. Земская Е.А., Китайгородская М.В., Ширяев Е.Н. Русская разговорная речь: Общие вопросы. Словообразование. Синтаксис. М., 1981
5. Лаптева О.А. Живая русская речь с телеэкрана. - М., 2000.
6. Русская грамматика т.1, М., Наука, 1980.

И.Е. Апушкина, А.В. Венцов,** Н.А. Слепокурова****

О фонетическом слове¹

Санкт-Петербургский государственный университет,

г. Санкт-Петербург, Россия

tchiric@mail.ru, **av.ventsov@mail.ru, *nataliars@inbox.ru*

Термин «фонетическое слово» (далее — ФС) давно и хорошо известен в лингвистике. Традиционное его понимание отражено в словаре О.С. Ахмановой [Ахманова 1966: 497]: «самостоятельное слово вместе с примыкающими к нему безударными служебными словами и частицами». Считается, что именно из ФС складываются синтагмы, являющиеся, в свою очередь, единицами не только акустико-фонетического, но и грамматико-смыслового членения речевого потока. Есть основания полагать, что ФС является одной из фонетических универсалий: «закономерности формирования фонетических слов существенно отличаются в разных языках. Однако во всех языках такая специфическая единица, как фонетическое слово, вызывается к жизни сходными причинами. Биологические ограничения на речь требуют ее ритмизованности, в то же время когнитивные нужды «подталкивают» к установлению корреляции между единицами ритма и единицами языка, представленными в речи. Образование фонетических слов и есть результат своего рода компромисса между двумя требованиями: возможность перераспределения лексического материала между фонетическими словами обеспечивает наличие одного из наиболее важных ритмов — словесного, а знание закономерностей распределения позволяет установить (вероятностные) корреляции между членением на ритмические и смысловые единицы текста» [Касевич и др. 1990: 225]. Более того, высказывалось предположение, что именно ФС могло бы быть единицей перцептивного словаря [Венцов, Касевич 2003: 214].

Зачастую понятие ФС признается как бы априорно известным и не подлежащим ни специальному толкованию, ни каким-либо исследованиям. А неясности в его отношении, тем не менее, существуют.

Не вызывает разногласий тот факт, что ФС не всегда совпадает со словом лексико-грамматическим, «определенным в соответствии с грамматическими критериями и интуицией носителей языка», как конкретизируется понятие слова в [Касевич 1977: 61]. Лингвистический энциклопедический словарь также указывает, что «фонетическое слово не совпадает с границами слова в его лексико-грамматическом понимании и может объединять ряд слов» [ЛЭС: 465].

Таким образом, ФС, в представлении лингвистов, может равняться по длине одному лексико-грамматическому слову, а может и состоять из нескольких слов. При этом не имеющие ударения слова в составе ФС называются проклитиками или энклитиками, в зависимости от того, с какой стороны они примыкают к слову с ударным слогом. Однако единого мнения по поводу того, какие слова могут становиться клитиками, среди лингвистов нет. Часть авторов причисляет к потенциальным клитикам и служебные, и знаменательные слова, остальные же, как правило, отказывают знаменательным словам в способности терять ударение в составе ФС.

К последним относится и О.С.Ахманова: в ее словаре указано, что самостоятельное слово сочетается с «примыкающими к нему безударными служебными словами и частицами» [Ахманова 1966: 497]. Л.Р. Зиндер также пишет о слиянии со значимым словом «служебных слов, лишенных в предложении самостоятельного ударения» [Зиндер 1960: 250]. В.Б. Касевич в [Касевич и др. 1990: 47] высказывается более осторожно: «Фонетическое слово в языках типа русского состоит обычно из полнозначной словоформы плюс служебные

¹ Работа выполнялась в рамках проекта «Русский язык и современная Россия» при поддержке федеральной целевой программы «Русский язык».

слова-клитки, которые функционально (или во всяком случае семантически) близки флексиям, они принадлежат слову не в меньшей степени, чем окончания». В других же работах, напротив, указывается, что «безударным в таком случае может оказаться как служебное, так и знаменательное слово» [Бондарко 1998: 217]. Л.В. Златоустова тоже говорит о сочетании «знаменательного и служебного (реже — двух знаменательных) слов, объединенных одним словесным ударением» [Златоустова 1977: 63]. И действительно, если бы знаменательные слова никогда не теряли своего ударения, было бы очевидно, что в каждом ФС может быть только одно знаменательное. Однако то, что знаменательное слово может выступать в составе ФС в качестве проклитики, иллюстрируется на простых и всем хорошо известных примерах: *за́ городом, под́ гору*, где ударение сохраняет предлог, а существительное оказывается энклитикой. В таком случае встает вопрос: знаменательное слово способно утрачивать свое ударение в рамках ФС только в сочетании со служебным словом? Возможно ли сочетание двух знаменательных слов, при котором одно из них теряет свое ударение, примыкая к другому?

Еще один немаловажный вопрос можно сформулировать так: всегда ли служебное слово примыкает к знаменательному? Может ли ФС не иметь в своем составе знаменательного слова? Большинство исследователей либо игнорируют этот вопрос, либо отвечают на него отрицательно. Однако некоторые все же рассматривают такую возможность: по мнению Т. Алексиевой, ФС «может состоять из одного или нескольких слов (полнозначных или служебных)» [Алексиева 2004: 168].

Трудности также возникают при попытке выделить основные критерии, условия образования ФС. Традиционно в качестве главного критерия выделяется ударение, тем более что «основная функция словесного ударения — фонетическое объединение слова в одно целое, обеспечение его цельноформленности. Эта функция осуществляется путем выделения одного из слогов в слове — ударного и подчинения ему остальных, безударных» [Бондарко 1998: 216]. Л.Р. Зиндер отмечал, говоря о ФС, что «точнее был бы термин «акцентная группа», которым подчеркивалась бы роль ударения в образовании подобных групп» [Зиндер 1960: 251]. Н.В. Богданова и Т. Алексиева тоже приписывают ударению ведущую роль в организации единства ФС [Богданова 1988; Алексиева 2004].

Тем не менее, ударение — не единственный критерий. В.Б. Касевич весьма неопределенно говорит о том, что отрезок текста, являющийся ФС, выделяется «по каким-либо фонетическим признакам, например объединяемый одним ударением, сферой действия сингармонизма и т. п.» [Касевич 1977: 61]. Т. Алексиева пишет, что «каждое фонетическое слово обладает рядом просодических признаков (длительностью, интенсивностью, частотой основного тона), в совокупности своей обеспечивающих его вычленимость в речи» [Алексиева 2004: 176]. Эти данные могут оказаться полезными, поскольку при решении конкретных прикладных задач опора исключительно на критерий ударения может оказаться несостоятельной.

Все вышеприведенные свойства и признаки ФС, однако, были обнаружены либо путем интроспекции, либо, в лучшем случае, путем изучения лабораторных образцов «правильной» речи. Когда же доступной для фонетического анализа сделалась звучащая естественная речь, стало очевидно, что подлинная языковая реальность во многом не совпадает с теоретическими представлениями о ней. Это касается и понятия ФС. Так, Н.В. Богданова пишет: «Анализ текстового материала позволил установить, что основной особенностью функционирования фонетического слова в спонтанной речи следует признать разрушение нормативного, литературного фонетического слова, осуществляемое двумя путями. Во-первых, в разговорных текстах наблюдается тенденция ударного произнесения служебных слов, выступающих в кодифицированном литературном языке, как правило, только в качестве проклитик или энклитик (*вот тут, под ураганым*)... Во-вторых, в речевое фонетическое слово на правах проклитик или энклитик часто включаются полнозначные, иногда длинные слова (*дали про-ект*)» [Богданова 1988: 108].

Из этого следует, что понятие ФС действительно нуждается в ревизии или, по меньшей мере, в уточнении. Задачей данного исследования было рассмотрение реального функциони-

рования ФС в потоке спонтанной устной речи. Образцом для исследования было выбрано ФС, состоящее из сочетания предлога и существительного. Такой тип ФС оказался в центре внимания с учетом критерия частотности и продуктивности. По свидетельству Т. Алексиевой [Алексиева 2004: 179], «при объединении словоформ в единое целое наиболее продуктивным способом следует признать проклитический». Об этом же говорят и данные Н.В. Богдановой, которая пишет: «Изучая способы образования текстовых фонетических слов и определяя количественную характеристику их продуктивности в экспериментальном материале, мы получили следующую картину. Во всех текстах наиболее продуктивным оказался проклитический способ образования». (92% и 91% в текстах КЛЯ и 75% и 57% в текстах разговорной речи, где наиболее частотными проклитиками являются предлоги: 45% и 74% в текстах КЛЯ и 47% и 40% в текстах разговорной речи) [Богданова 1988: 109].

В качестве предлогов рассматривались только простые первообразные предлоги из списка, данного в Академической грамматике [Русская грамматика 1982: 707]. Интересно заметить, что этот список практически полностью совпадает с подмножеством абсолютных проклитик современного русского языка, которое составляют предлоги, по данным [Лефельдт 2006: 132].

Как было показано выше, для решения вопроса о том, одно или два ФС представлено в тексте, можно прибегнуть к оценке разных параметров: ударения (сравнивать длительность ударной гласной или вокализованного сегмента ударного и потенциального ударного слогов), сегментной редукции (как явления, связанного с предположительно качественной природой русского ударения), мелодических характеристик. Однако здесь уместно будет вспомнить второе определение ФС, данное в словаре [Ахманова 1966: 497], где сказано, что ФС — это «отрезок речи, не имеющий в своем составе открытого соединения», т.е. не содержащий в своем составе паузы. В.Б.Касевич тоже говорит о паузе как о фонологическом средстве, пограничном сигнале, отсутствие которого приводит к нейтрализации противопоставления двух словосочетаний. Однако, что важно, он упоминает, что «гортанная смычка или гласный с гортанносмычным началом обеспечивают словоделение, не будучи фонемами». Последний есть основание рассматривать как постпаузальный аллофон фонемы — следовательно, его наличие «свидетельствует о присутствии паузы, хотя таковая и не представлена акустически» [Касевич 2002: 9].

Материалом для исследования послужили аудиозаписи телепрограммы «Культурная революция» общей продолжительностью в 3 часа 22 минуты (~202 минуты). Экспериментатор, пользуясь программным пакетом Cool Edit Pro, осуществлял сплошное прослушивание аудиозаписей в поисках пауз внутри словосочетаний предлогов с существительными. В случае обнаружения такой паузы ее длительность измерялась с помощью программных средств и фиксировалась в тексте расшифровки (в квадратных скобках указывалось время начала и конца паузы, ее длительность и комментарии). При решении вопроса о минимальной длительности паузы приходилось полагаться на осциллограмму и слуховое ощущение исследователя, так что результат следует признать в какой-то мере субъективным.

Таким образом был получен список словосочетаний, состоящих из существительного и предлога, разделенных паузой.

Следующим этапом исследования было выяснение того, какую долю из общего числа предлогов, встретившихся в исследуемом аудиотексте в составе словосочетаний с существительными, составляют разделенные паузой словосочетания. Количество предлогов определялось по текстовой расшифровке указанных аудиозаписей с помощью специально для этой цели написанной программы. Оно приведено ниже.

БЕЗ — 5	ОБ — 2
ВО — 2	ОТ — 14
ДЛЯ — 10	ПЕРЕД — 1
ДО — 6	ПО — 12
ЗА — 13	ПРИ — 1

ИЗ — 12	ПРО — 7
ИЗ-ЗА — 5	РАДИ — 1
КРОМЕ — 1	СО — 5
НА — 27	У — 7
О — 13	ЧЕРЕЗ — 1

Всего в тексте, таким образом, встретилось 145 словосочетаний существительных с вышеперечисленными 20 предлогами (из 38, зафиксированных в Русской грамматике-1982).

Что касается словосочетаний предлог–существительное, разделенных паузами, то их было обнаружено 11:

до [157 мс] *появления*
про [98 мс] *магов*
от [247 мс] *передозировки*
до [598 мс] *гостиницы*
о [366 мс] *запрещении*
от [162 мс] *кино*
без [268 мс] *воспитания*
на [604 мс] *Турин*
из [681 мс] *спортсменов*
для [142 мс] *подъема*
по [323 мс] *объемам*

При этом в случаях 4, 6, 7, 9, 11 пауза была хезитационной, заполненной. К последним, кажется, стоит добавить еще один: *спасибо за* [241] *так сказать идею*. По крайней мере, здесь предлог однозначно отделен от существительного паузой (хезитационной) и не составляет с ним единого фонетического слова.

Встретилось также еще несколько употреблений предлогов с существительными, по следующей схеме: предлог, пауза, повторение предлога, существительное. При этом второй предлог с существительным явно составляли нераздельное целое.

по [627 мс] *по масштабу*
по [148 мс] *по мере*

К сожалению, второй случай не подходит под условие исследования, потому что в данном контексте это сочетание является составным именным предлогом.

В процессе обработки материала было также зафиксировано 9 случаев паузации между предлогом и другими частями речи, два из которых были представлены гортанной смычкой в начале существительного, а также один случай паузы после составного именного предлога *в отличие от*. Стоит упомянуть также о двух случаях, когда пауза после предлога заканчивала реплику говорящего:

борца за [157 мс]
двадцатый век на [778]

В последнем случае большая длительность паузы объясняется сменой говорящего, это была пауза между двумя репликами разных людей.

Итак, всего было зафиксировано 12 случаев безусловного наличия паузы между предлогом и существительным. Из них в 6 паузы были заполненными, хезитационными, в остальных 6 случаях паузе соответствовало отсутствие речевого сигнала. Обобщение данных представлено в табл. 1.

Таблица 1.

Предлог	Употребление с паузой перед существительным	Употребление в тексте перед существительным
БЕЗ	1 (20%)	5
ДЛЯ	1 (10%)	10
ДО	2 (33,3%)	6
ЗА	1 (7,7%)	13
ИЗ	1 (8,3%)	12
НА	1 (3,7%)	27
О	1 (7,7%)	13
ОТ	2 (14,3%)	14
ПО	1 (8,3%)	12
ПРО	1 (14,3%)	7

Рассматривая полученные результаты, можно с уверенностью говорить о том, что данные, полученные в ходе исследования, не согласуются с традиционным пониманием ФС: служебные слова, которым теория предписывает потерю ударения и вхождение в состав фонетического слова вместе с последующим существительным, в ряде случаев оказываются отделенными от существительного паузой. Кроме того, очевидно, что частота употребления предлога в качестве отдельного фонетического слова, по видимости, никак не коррелирует с его частотностью в данном тексте.

Данные эксперимента Н.В. Богдановой, упоминаемого выше, позволяют говорить о том, что «наблюдается тенденция ударного произнесения служебных слов, выступающих в кодифицированном литературном языке, как правило, только в качестве проклитик или энклитик» [Богданова 1988: 108]. В нашем случае, при наличии паузы, мы можем с уверенностью говорить лишь об обнаружении границы между фонетическими словами. И из того, что свойством каждого фонетического слова является наличие ударения, было бы логично ожидать, что предлог, отделенный от существительного паузой, окажется ударным. Однако вопрос о подтверждении или опровержении этого факта остается открытым: он требует проведения дополнительного инструментального анализа и перцептивного эксперимента. По субъективному впечатлению одного из исследователей, которое, правда, не может считаться показательным, большинство предлогов казались на слух безударными, сильно редуцированными, а часть (например, *про* из *про магов*) воспринималась как лишенные ударения, но качественно нередуцированные. Таким образом, вопрос о длительности и сегментной редукции предлога, составляющегося отдельное ФС, подлежит дальнейшему исследованию.

Выводы

1. Традиционное понятие ФС подлежит пересмотру с привлечением данных устной спонтанной речи.
2. Проблемными для ФС остаются вопрос о его границах, составе, месте ударения и клитиках внутри ФС, о состоятельности принципа «сколько ударений, столько фонетических слов».
3. По данным проведенного исследования, предлог и существительное в спонтанной речи могут не объединяться в одно фонетическое слово, оказываясь разделенными паузой. Таким образом:
 - а. фонетическое слово может состоять из предлога, вопрос об ударности которого остается открытым;
 - б. сочетание служебного слова с существительным в рамках одного фонетического слова и, тем более, потеря предлогом своего ударения, — не абсолютное правило.
4. Вопросы о длительности ударных гласных или вокализованных сегментов ударных или потенциально ударных слогов в составе фонетического слова, о сегментной редукции и мелодических характеристиках в рамках фонетического слова, а также об услови-

ях формирования фонетического слова из нескольких лексико-грамматических слов подлежат дальнейшему исследованию.

5. Полученные экспериментальные делают весьма проблематичной гипотезу о том, что ФС может быть единицей перцептивного словаря.

Литература

1. Алексиева Т. Фонетическое слово как единица анализа звучащей речи (на материале русского и болгарского языков) // Язык и речь: проблемы и решения: Сб. науч. тр. к юбилею профессора Л.В. Златоустовой / Под ред. Г.Е. Кедровой, В.В. Потапова. М., 2004. С. 168-181.
2. Ахманова О.С. Словарь лингвистических терминов. М., 1966.
3. Богданова Н.В. Фонетическое слово как единица описания спонтанного речевого текста // Фонетика спонтанной речи / Под редакцией Светозаровой Н.Д. Л., 1988. С. 105-111.
4. Бондарко Л.В. Фонетика современного русского языка. СПб., 1998.
5. Венцов А.В., В.Б.Касевич. Проблемы восприятия речи. М., 2003.
6. Зиндер Л.Р. Общая фонетика. Л., 1960.
7. Златоустова Л.В. Звуковые типы русской речи // Проблемы теоретической и экспериментальной лингвистики. М., 1977. С. 146-160.
8. Касевич В.Б. Элементы общей лингвистики. М., 1977.
9. Касевич В.Б., Е.М. Шабельникова, В.В. Рыбин. Ударение и тон в языке и речевой деятельности. Л., 1990.
10. Касевич В.Б. Еще о понятии фонетического слова // Проблемы фонетики IV. М., 2002. С. 7-12.
11. Лефельдт В. Акцент и ударение в русском языке. М., 2006.
12. ЛЭС: Лингвистический энциклопедический словарь / Под ред. В. Н. Ярцевой. М., 1990.
13. Русская грамматика. Т. I: Фонетика. Фонология. Ударение. Интонация. Словообразование. Морфология / Гл. ред. Н.Ю. Шведова. М., 1982.

Г.В. Дорохина*, Т.В. Ермоленко**

Модели фонемного распознавания речи, явно использующие свойства языка и речи

*Институт проблем искусственного интеллекта МОН и НАН Украины,
г. Донецк, Украина*

* sgv@iai.donetsk.ua, ** etv@iai.donetsk.ua

1. Актуальность

Сложную, не решенную на сегодняшний день проблему распознавания речи считают актуальной с 50-х годов прошлого столетия. Изучение речи как многоуровневой системы знаков и способов взаимодействия между знаками различных уровней тесно связано с моделированием интеллектуальной деятельности человека и требует привлечения средств компьютерной лингвистики и других направлений искусственного интеллекта, а успех в этой области будет способствовать развитию представлений о речевой деятельности человека, интеллектуализации компьютерных интерфейсов, ускорению и упрощению процесса человеко-машинного общения.

2. Анализ публикаций

В последнее время в области речевых технологий можно отметить использование статистических методов описания речи на акустическом, лексическом и более высоких уровнях. Это оправдано, если в основании лежит адекватная модель объекта. Кроме того, неявное использование свойств объекта в узких практических целях не развивает науку о самом объекте.

Методы распознавания речи можно классифицировать их следующим образом.

1. Статистические.

- a. Единицы обучения и распознавания **совпадают**. К ним относятся: метод динамической временной деформации (ДВД); метод нейронных сетей.
- b. Единицы обучения и распознавания **не совпадают**. К этой группе методов принадлежат: метод скрытых Марковских моделей (СММ); метод КДП, основанный на композиции (К) и динамическом программировании (ДП)[1].

2. Использующие явные знания о языке и речи.

- a. Методы акустическо-фонетического подхода, в явном виде использующие данные акустической фонетики. Они предполагают явную сегментацию речевого сигнала (РС) и классификацию полученных сегментов, а для определения последовательности произнесённых слов используют статистическую модель языка.
- b. Методы искусственного интеллекта (ИИ). Базируются на использовании знаний акустического, лексического синтаксического и семантического уровней.

В последнее время в области речевых технологий характерно использование СММ как основного метода описания речи на акустическом, лексическом и более высоких уровнях. Это позволяет решать локальные задачи. Однако сильная зависимость этого метода от обучающей выборки, а также высокая продуктивность и нестатистическая природа высших уровней языка [2] сильно ограничивают его применение для распознавания естественной речи. Методы акустическо-фонетического направления, явно использующие данные акустической фонетики, на последующих уровнях обработки также опираются на статическую модель языка.

По мнению авторов, применение именно статистической модели языка при распознавании слитной речи вызвано следующими причинами.

1. В распознавании слитной речи не указано каких-либо ограничений на длину распознаваемого фрагмента.
2. Отсутствуют модели, в аналитическом виде описывающие синтаксические и семантические закономерности речи без ограничения длины анализируемого фрагмента. При наличии таких моделей без ограничения длины анализируемого фрагмента модель становится «вычислительно неподъёмной» для любой системы распознавания.
3. Как отмечено в работе [1], системы распознавания речи должны быть построены на принципах ИМЗ иерархической (И) обработки данных и принятия многозначных решений (МЗ), что приводит к необходимости хранения и обработки составных гипотез. При невозможности хранить и обрабатывать большое количество гипотез исследователи вынуждены были выбирать небольшое количество наиболее вероятных вместо анализа всех возможных.

Использование способа представления составных гипотез [3, 4], который обеспечивает их компактное хранение и скоростную обработку, позволит на каждом этапе распознавания исключать из рассмотрения только недопустимые в языке гипотезы и выполнять анализ РС в соответствии с используемыми схемами обработки речи (рис 1. а-в), или заменять их более адекватными.

Целью работы является разработка моделей фонемного распознавания речи на основе явного использования свойств языка и речи. Данная цель определила задачи исследования:

- дать характеристику русской речи как объекту исследования;
- с учётом этих характеристик разработать схему обработки речи;
- согласно предложенной схеме разработать модели фонемного распознавания на основе методов направления искусственного интеллекта;
- проанализировать области применения разработанных моделей и способы их реализации.

3. Характеристика русской речи как объекта исследования

В работе [2] выделены такие «системообразующие факторы речевой системы», как:

1. Продуктивность — возможность продуцировать сколь угодно большое количество информационных сообщений, обладающих разным смыслом.
2. Помехозащищённость — способность «обеспечить сохранение смысловой информации при различных вариантах акустических помех и искажений» даже «при различных патологических или ситуационных нарушениях процессов речеобразования и речевосприятия. Обеспечивается путём использования «нескольких параллельно работающих способов выделения одних и тех же элементов речевого сигнала на базе анализа акустического сигнала» и параллельного независимого «использования сегментного (фонемного) и целостного восприятия слов в потоке речи».

В работе [5] речь рассматривается как помехоустойчивый звуковой код. В ней высказано предположение, что речь — это блочный код. Кодирование производится не за счёт *предъявления* в частотной области сигналов определённой интенсивности, а за счёт *изменения* спектральных характеристик и интенсивности. Декодирование возможно только при рассмотрении некоторого *блока* информации. Это позволит при распознавании *отказаться от абсолютных порогов* и определять их исходя из полученного блока речи (БР). Таким блоком речи является фрагмент речи, отделённый от остальной речи межфразовыми паузами (изолированное слово или слитно произносимая фраза). Т.е. в основе распознавания слитной речи должно лежать распознавание БР — фрагментов конечной длины (т.к. не содержат интонационных и дыхательных пауз) с потенциально формализуемой синтаксической структурой. Далее последовательность слов, соответствующих слитно произносимой фразе, будем

называть фразой. В работе [6] фразы предложено считать простыми распространёнными несложнёнными предложениями.

По мнению авторов используемые в современных системах распознавания схемы обработки речи (рис. 1 а-в) вступают в некоторое противоречие с понятием «фонема», которое даёт фонетика. Фонемы — это смысловозначительные единицы, а не единицы различения звуков (соответствуют очень размытым в пространстве акустических признаков классам). Представление речи как последовательности фонем следует рассматривать уже после определения множества гипотез слова или фразы.

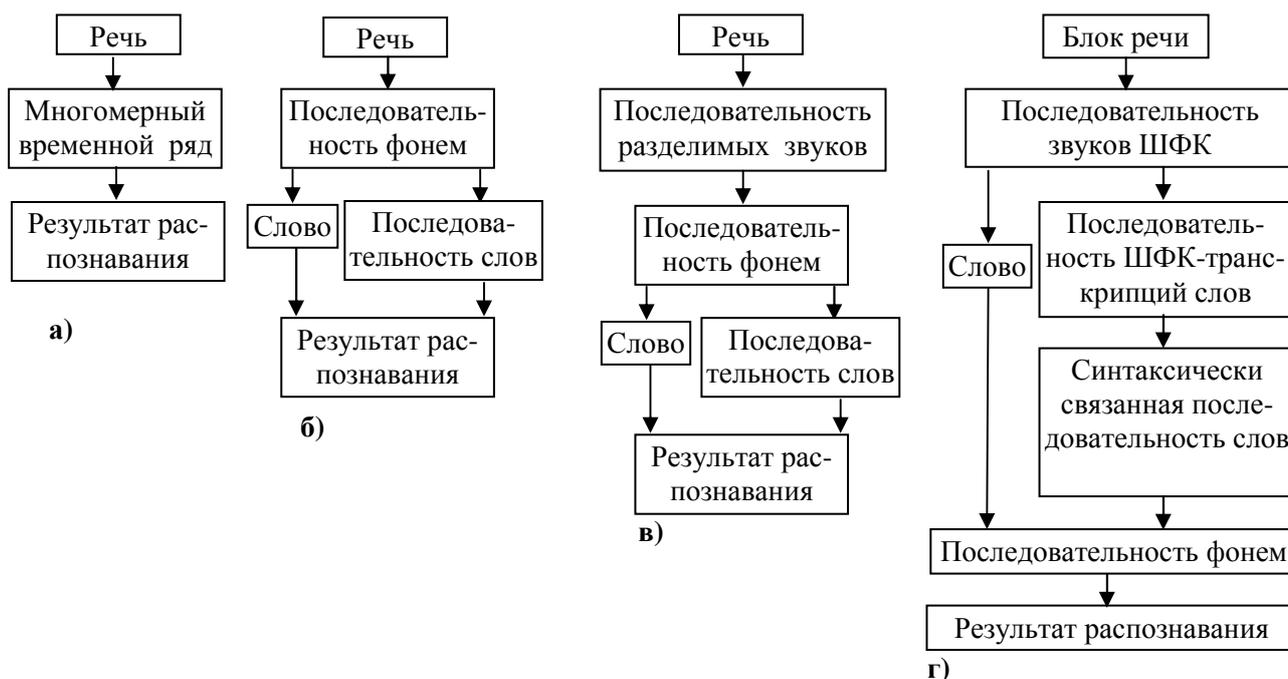


Рис. 1 Схемы обработки речи различными методами: а) ДВД, б) СММ, в) методами акустико-фонетического направления, г) предложенная схема обработки блока речи.

Хорошо же разделимыми в пространстве акустических признаков являются широкие фонетические классы (ШФК) звуков образованные различными источниками звука: невокализованные смычки ([п, т, к]); невокализованные фрикативные ([ш, щ, ч, с, ц, ф, х]); вокализованные фрикативные ([ж, з, д̄ж, д̄з]); остальные вокализованные согласные и гласные. Искажения РС, приводящие к ошибкам в различении ШФК-классов, сделают этот РС нераспознаваемым и для человека. Следовательно, выдвижение или исключение гипотез о произнесении необходимо выполнять по результатам анализа последовательности ШФК звуков РС.

Исходя из вышесказанного, можно заключить, что схема обработки речи должна иметь вид (рис. 1г). За счёт анализа всех допустимых кандидатов такая схема обеспечивает параллельное независимое использование сегментного (фонемного) и целостного восприятия потока речи. Ограничение на длину распознаваемого БР позволяет предположить формализуемость фразы как последовательности слов, а анализ речи блоками может содействовать переходу в распознавании речи от использования абсолютных порогов к относительным, вычисляемым по БР.

4. Модели пофонемного распознавания речи

Рассматривая слитную речь как последовательность БР, в задаче распознавания получаем две подзадачи: распознавания БР и выбор последовательности БР. Выбор после-

довательности БР — задача семантического уровня, в явном виде неразрешимая нами на данном этапе. Рассмотрим подзадачу распознавания БР.

Ориентируясь на разработку моделей и алгоритмов фонемного распознавания речи, явно использующих свойства языка и речи, попадаем в область ИИ, где используют методы интеграции источников знаний «снизу-вверх» и «сверху-вниз». С учётом предложенной схемы обработки блока речи разработаны модели фонемного распознавания (рис. 2).

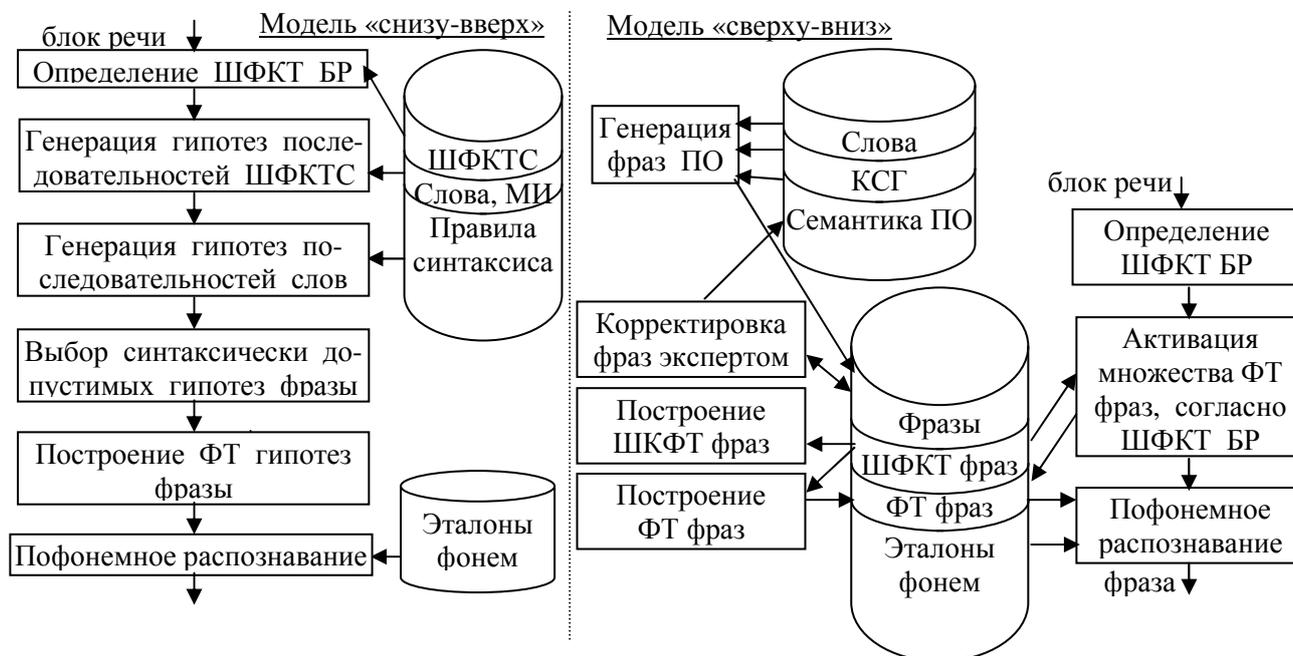


Рис. 2 Разработанные модели фонемного распознавания при методах интеграции источников знаний «снизу-вверх» и «сверху-вниз».

На рис.2 использованы следующие обозначения: БР — блок речи; ШФКТ — транскрипция согласно широкой фонетической классификации; ШФКТС - транскрипция слова согласно широкой фонетической классификации; МИ — морфологическая информация о слове (часть речи, род, число и пр.); ФТ — фонетическая транскрипция; ПО — предметная область; КСГ — контекстно-свободная грамматика, описывающая синтаксис языка.

Данные модели в качестве явно заданных закономерностей речи используют для ограничения количества гипотез фраз ШФК-транскрипции БР. Закономерности языка отражают правила синтаксиса, которые описывают синтаксис простых распространённых неосложнённых предложений, а закономерности предметной области — соотношение понятий, выраженное в возможности участия групп слов в определенных видах синтаксической связи.

Распознавание речи при интеграции источников знаний «снизу-вверх» может применяться в задачах диктовки, стенографирования, распознавания спонтанной речи — т.е. в случае, когда трудно заранее оценить перечень возможных фраз и речь нужно фиксировать, а не ставить ей в соответствие некоторое действие. Система, моделирующая этапы «Генерация гипотез последовательностей ШФКТС», «Генерация гипотез последовательностей слов», «Выбор синтаксически допустимых гипотез фразы», описана в работе [6]. При этом множества гипотез (последовательностей ШФКТС, последовательностей слов, фразы) генерируются, анализируются и хранятся в явном виде, благодаря способу представления составных гипотез, изложенному в работе [3].

Системы распознавания команд и вопросно-ответные системы предполагают выполнение определённого действия как реакции на результат распознавания. Для обеспечения устойчивой работы системы перечень фраз и действий целесообразно определять до этапа эксплуатации. В таком случае логично применение метода интеграции знаний «сверху-вниз».

По сравнению с распознаванием слитно произносимых фраз согласно модели «снизу-вверх», синтезирующей множество гипотез фразы и анализирующей эти гипотезы, в данной

выполняется поиск множества гипотез распознаваемой фразы, что требует значительно меньших вычислительных затрат.

Отметим, что для системы распознавания в этом случае не имеет значения распознаёт она слово или фразу — и то, и другое известно системе и представляется последовательностью фонем. Т.е. для этапа «Пофонемное распознавание» (как для модели «снизу-вверх», так и для модели «сверху-вниз») можем использовать алгоритмы пофонемного распознавания изолированных слов [7]. Фонетическую транскрипцию каждой фразы можем получить с помощью транскриптора, использованного в работе [8], а ШФК-транскрипцию фразы по её написанию — с учётом [6]. Формирование структур словаря фраз, обеспечивающих их скоростной поиск выполним согласно [4]. Получение ШФК-транскрипции БР можно выполнять опираясь на работы [9-10]; поиск в словаре фраз множества гипотез фразы, соответствующих ШФК-транскрипции фрагмента речи, обеспечивается сформированными на подготовительном этапе структурами словаря фраз с учётом [4].

Следовательно, в указанной модели «снизу-вверх» непроработанным к настоящему моменту остался вопрос синтеза фраз предметной области, чему посвятим последующие работы.

5. Заключение

В работе рассмотрены методы распознавания и используемые в них схемы обработки речи. Указаны возможные причины использования при распознавании речи исключительно статистических моделей речи и показаны недостатки такого подхода.

Дана характеристика речи как объекту исследования, на основе которой предложено рассматривать слитную речь как последовательность БР и сформирована схема обработки блоков речи.

В соответствии с данной схемой разработаны модели пофонемного распознавания, предполагающие интеграцию источников знаний «снизу-вверх» и «сверху-вниз», определены задачи распознавания, в которых они могут быть применены, указаны способы реализации элементов системы пофонемного распознавания,

К отличительным особенностям работы относятся явное задание и использование в модели закономерностей речи и языка; хранение и анализ составных гипотез, представленных согласно [4]; исключение из рассмотрения только недопустимых гипотез, а не анализ наиболее вероятных.

Литература

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – Киев, 1987.
2. В.И.Галунов, А.Н.Соловьев Современные проблемы в области распознавания речи.// Информационные технологии и вычислительные системы, № 2, 2004г.
3. Патент України № 78806 «Пристрій для збереження і пошуку рядкових величин та спосіб збереження і пошуку рядкових величин» Власник: Інститут проблем штучного інтелекту, винахідник Дорохіна Г.В. // Промислова власність. Бюл. № 5, 25.04.2007.
4. Дорохіна Г.В. Применение изобретения «Устройство для хранения и поиска строковых величин» // Винахідник і раціоналізатор.- 2007. – № 11. – с. 12-13.
5. Саввина Г.В. Речь как звуковой код // Труды конф. «Информационные технологии в социологии, экономике, образовании и бизнесе» Ялта-Гурзуф (Украина, Крым). – 30 сентября – 9 октября 2003 г. – Запорожье. – 2003. – С. 49-51.
6. Дорохіна Г.В. Ограничение количества гипотез фразы при распознавании слитной речи // Известия ТРТУ – 2005. – № 10. – С. 54-60.
7. Козлов А.В., Саввина Г.В., Шелепов В.Ю. Система пофонемного распознавания отдельно произносимых слов// Искусственный интеллект. – 2003. – № 1. – С. 156-165.
8. Грабовая В.А. , Федоров Е.Е., Шелепов В.Ю. О системе компьютерного распознавания русской речи с автоматическим построением эталонов // Искусственный интеллект. - 2000. - №.1. - С.76-81.
9. Ниценко А.В., Шелепов В.Ю. Алгоритмы пофонемного распознавания слов наперед заданного словаря // Искусственный интеллект. – 2004. – № 4. – с. 633-639.
10. Ермоленко Т.В. Разработка системы распознавания изолированных слов русского языка на основе вейвлет-анализа // Искусственный интеллект. – 2005. – № 4. – С. 595-601.

П.А. Скрелин, В.В. Евдокимова***

Вариативность реализаций гласных фонем в спонтанной речи и чтении

*Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия*

**skrelin@phonetics.pu.ru, **postmaster@phonetics.pu.ru*

Одним из перспективных направлений развития фонетики является изучение живой разговорной — спонтанной — речи (СР) во всем ее многообразии. Выделение фонетических особенностей СР, их формализация и классификация дает возможность поднять на качественно новый уровень решение ряда задач анализа и синтеза речи. Работы фонетистов подтвердили высокую вариативность СР, значительное влияние экстралингвистических факторов, позволили описать закономерности трансформации отдельных фонем и фонетических слов [1, 2, 3]. Существенной особенностью СР является изменение темпа (повышенный или пониженный темп) и, соответственно, скорость следования реализаций фонем и фонетических слов. Общие экстралингвистические условия протекания разговорной речи приводят к сходству происходящих в разных языках процессов и явлений. Среди них отметим: выпадение гласных и согласных, различного рода ассимиляции как внутри, так и на стыках слов, приводящие наряду с выпадением конечных звуков и слогов, к стиранию словесных границ [2].

Структурная избыточность речевого сообщения дает возможность говорящему держаться в соответствии с принципом экономии усилий, введенным А. Мартине, на той границе уровня насыщенности информационного потока, которая обеспечивает восприятие слушателем смысла речевого сообщения [4]. Такой речи присущ высокий уровень возможных звуковых изменений, что предельно усложняет процесс автоматического анализа речевых сообщений.

Для автоматической интерпретации данных акустического, а в дальнейшем и статистического анализа материала необходимо установить фонемную принадлежность каждого звука на основании его акустических характеристик. В отличие от анализа речевого сигнала человеком, формальные процедуры анализа сигнала не могут использовать сведения, реализуемые на более высоких уровнях системы языка и известные человеку из знания широкого контекста и ситуации общения. Как известно, (см. например, [5]) на месте фонемы /a/ может быть реализован практически любой звук из таблицы кардинальных гласных. То же можно сказать и о любой другой гласной фонеме. Для объяснения этого факта можно использовать «стационарность второго порядка» [5], «мену» (или «замену» фонем), подчеркивающую, что это явление носит случайный характер, либо имеет невыясненную регулярность или систематичность (иначе можно было бы говорить о чередовании фонем). Разработчика в этом случае должны интересовать исключительно формальные признаки сегментных единиц, а точнее, реализованные в речевом сигнале дифференциальные признаки гласных фонем — ряд, подъем, огубленность — в той мере, в какой они реализованы акустически, аналогично идентификации согласных (см. [6]).

Целью данного исследования является анализ аллофонического варьирования гласных в разных типах связной речи — спонтанной речи и чтении — путем сопоставления разных видов транскрипции гласных звуков.

Звуковым материалом послужили записи спонтанных диалогов и чтения 10 дикторов (5 мужчин и 5 женщин). Для каждого диктора из спонтанного диалога было выбрано по 5 минут звучания. Из записи диалога были составлены тексты таким образом, чтобы синтаксическая структура и лексический состав по возможности сохранялись. Затем тексты, составленные из диалогов, были предложены для прочтения тем же дикторам. Далее проводилась сегментация речевого корпуса, включающего 2 вида речи, на различные по размеру единицы:

звуки, слоги, фонетические слова, интонационные единицы, фразы. Сегментация на звуки осуществлялась на основании реализованной звуковой цепочки, при этом не было цели обязательно «найти» столько же звуков, сколько должно было быть в соответствии с идеальной фонемной последовательностью.

Транскрипция осуществлялась последовательно, в несколько этапов [7, 8]. Сначала был создан уровень акустической транскрипции, на котором каждому элементу акустической последовательности, приписывалась метка, исходя из его перцептивных и акустических характеристик. Далее создавалась идеальная транскрипция, которая представляет собой фонемное представление теоретически ожидаемой структуры звуковой последовательности. Поскольку звуки могли изменяться и выпадать, количество элементов на уровне «идеальной» транскрипции и их качество не совпадали с транскрипцией на «акустическом уровне».

Чтобы сравнить два типа речи — спонтанную речь и чтение — представляется интересным проанализировать количество и закономерности аллофонического и фонемного варьирования гласных при произнесении в сравнении с теоретически ожидаемой фонемной структурой. Для этого было проведено сопоставление двух типов транскрипции — акустической и «идеальной».

На Рис. 1а и 1б — в этой таблице разделены неогубленные (1а) и огубленные (1б) гласные, которые встретились в исследуемом материале.

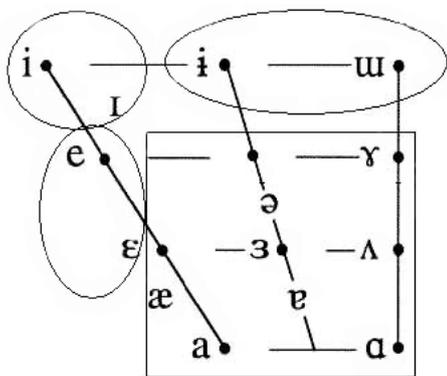


Рис. 1а. Фонемная принадлежность неогубленных гласных.

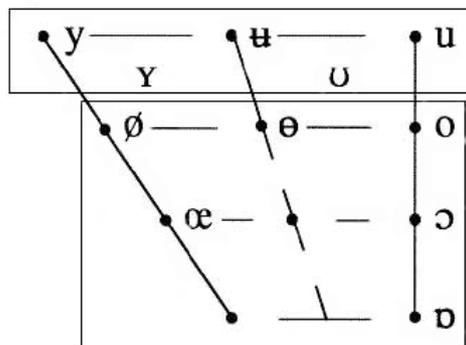


Рис. 1б. Фонемная принадлежность огубленных гласных.

При помощи принципа наименьшего расстояния от основного звукотипа гласные были объединены в зоны действия фонем. Установление фонемной принадлежности каждого гласного на основе формальных признаков позволило произвести анализ встречаемости фонем в спонтанной речи, оценить количество замен фонем, определить степень устойчивости фонем к заменам.

С помощью специально разработанных скриптов для программы обработки звуковых сигналов Praat [9] были получены данные по сравнению гласных «идеального» и акустического уровня разметки.

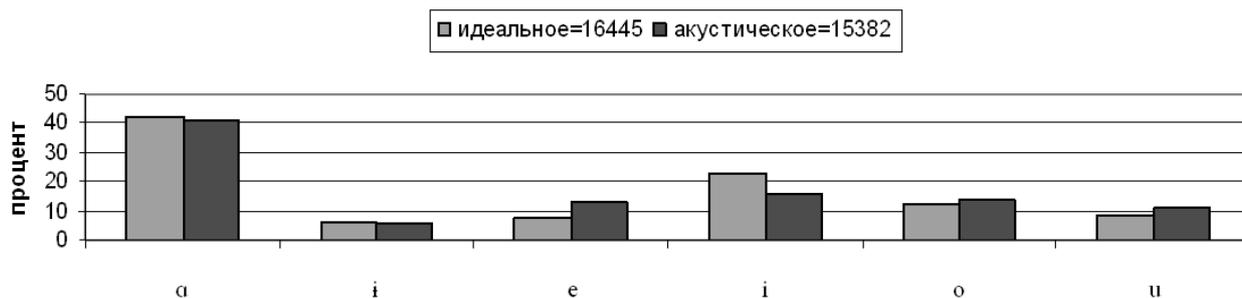


Рис. 2. Идеальное и реальное распределение гласных фонем.

Рис. 3. показывает сопоставление идеального и реального распределения фонем в материале. Распределение показывает, что больше всего «чужими» аллофонами заменяются фонема /i/, а довольно активными в этом процессе оказываются аллофоны /e/, /u/ и /o/. То, что число фонем превышает число аллофонов, объясняется тем фактом, что некоторые аллофоны были реализованы на месте последовательностей из двух и более фонем, включая согласные, т.е ряд фонем не получили физической реализации.

На Рис. 4 представлено соотношение гласных звуков, реализованных в исследованном материале.

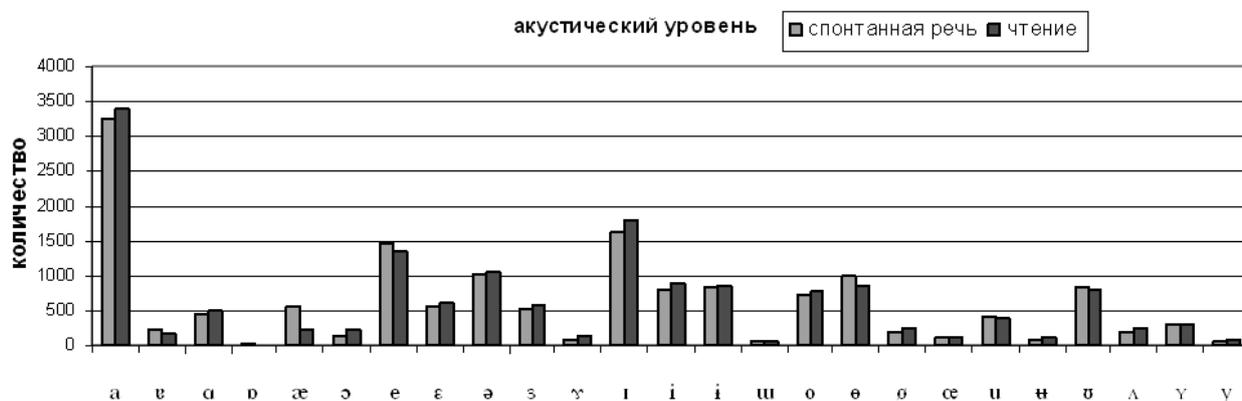


Рис. 3. Соотношение гласных звуков, реализованных в исследованном материале.

Рис. 5 – 10 показывают, какие звуки реализованы на месте фонем, формирующих идеальный фонемный состав словоформ.

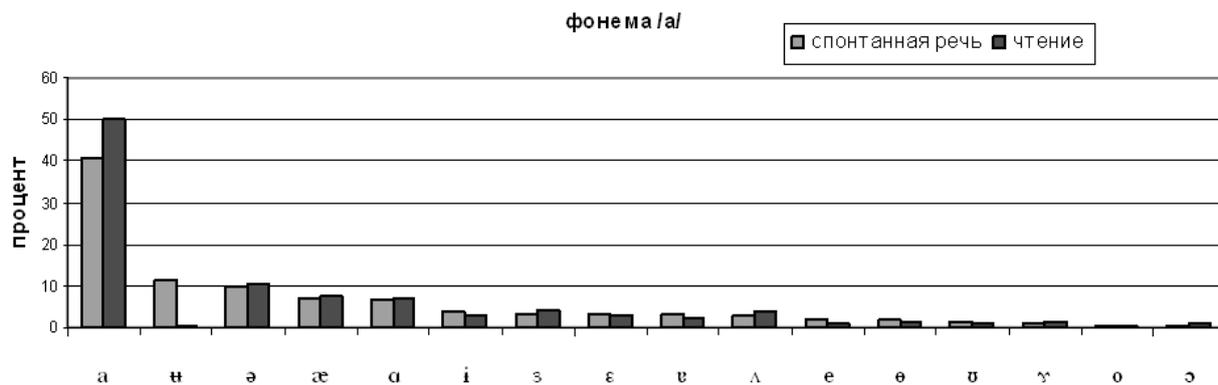


Рис.4. Соотношение звуков, реализованных на месте фонемы /a/ в спонтанной речи и при чтении.

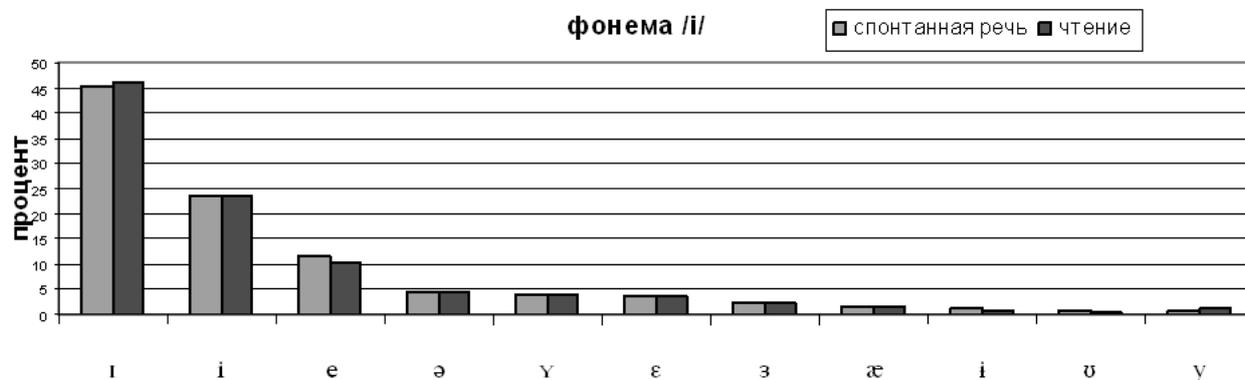


Рис.5. Соотношение звуков, реализованных на месте фонемы /i/ в спонтанной речи и при чтении.

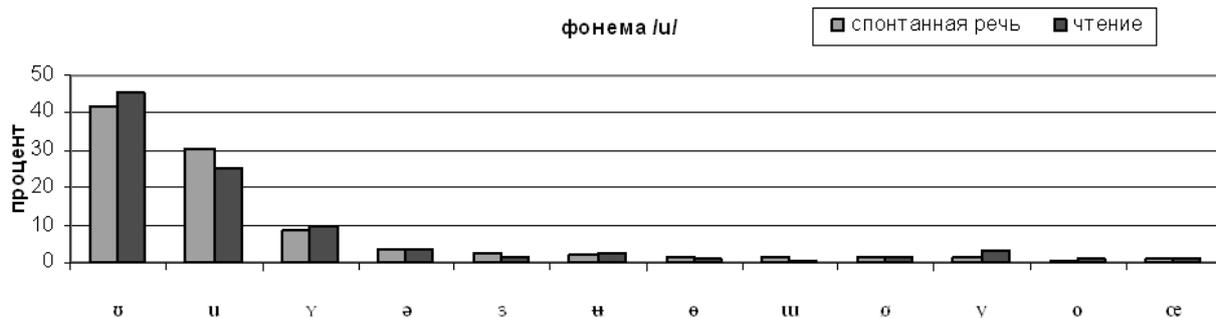


Рис.6. Соотношение звуков, реализованных на месте фонемы /u/ в спонтанной речи и при чтении.

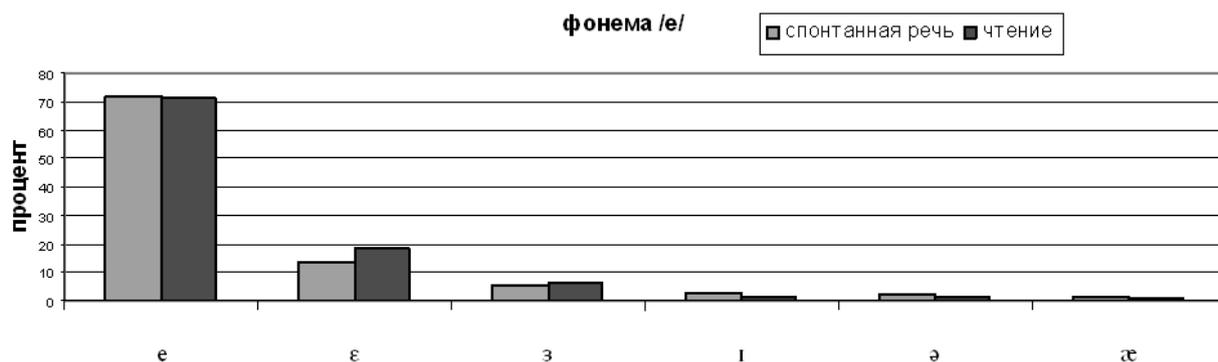


Рис.7. Соотношение звуков, реализованных на месте фонемы /e/ в спонтанной речи и при чтении.

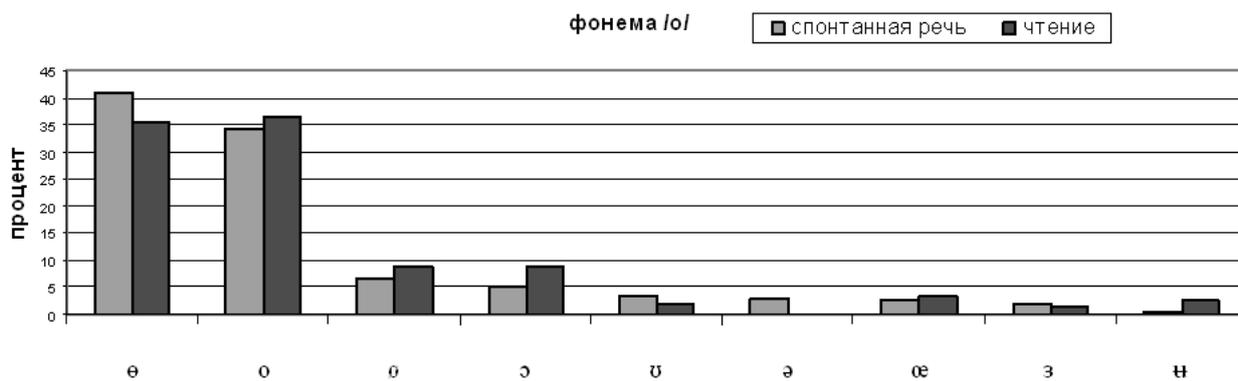


Рис.8. Соотношение звуков, реализованных на месте фонемы /o/ в спонтанной речи и при чтении.

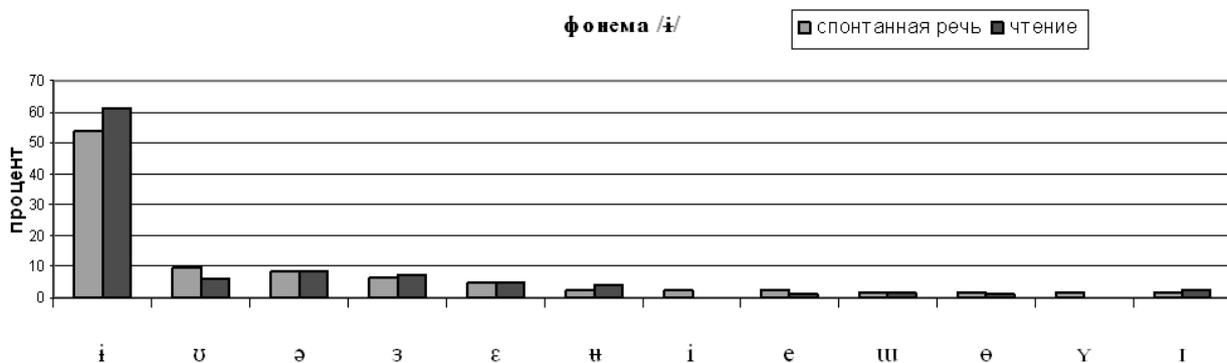


Рис.9. Соотношение звуков, реализованных на месте фонемы /i/ в спонтанной речи и при чтении.

Таблица 1. Процент аллофонов, реализованных на месте фонемы /a/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	74,1	87,4
і	3,8	3,2
е	5,2	4,0
і	0,3	0,0
о	3,6	3,6
u	13,0	1,8

Таблица 3. Процент аллофонов, реализованных на месте фонемы /ɨ/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	7,2	6,0
і	1,9	1,4
е	0,9	1,0
і	1,2	0,0
о	4,8	5,0
u	84,0	86,6

Таблица 5. Процент аллофонов, реализованных на месте фонемы /o/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	4,9	1,9
і	0,6	0,4
е	0,0	0,0
і	0,3	0,0
о	89,4	92,8
u	4,9	4,8

Таблица 2. Процент аллофонов, реализованных на месте фонемы /i/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	8,7	9,1
і	1,2	0,8
е	15,1	14,0
і	68,5	69,1
о	0,7	1,7
u	5,8	5,4

Таблица 4. Процент аллофонов, реализованных на месте фонемы /e/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	8,6	8,3
і	0,9	0,5
е	85,7	89,5
і	2,8	1,3
о	2,1	0,4
u	0	0,0

Таблица 6. Процент аллофонов, реализованных на месте фонемы /i/, сгруппированных по принадлежности к определенной фонеме

Фонема	Спонтанная речь	Чтение
а	16,7	16,3
і	55,4	63,2
е	7,2	6,0
і	4,3	2,6
о	2,0	1,9
u	14,3	10,0

Таблица 7. Процентное соотношение реализаций аллофонов, принадлежащих каждой фонеме и числа аллофонов, принадлежащих другим фонемам. Сравнение спонтанной речи и чтения.

фонема	Спонтанная речь	Чтение
	Количество реализаций «своих» аллофонов	Количество реализаций «своих» аллофонов
а	74,1	87,4
і	54,3	63
е	85,6	89,5
и	68,5	69
о	89,4	92,8
u	84	86,6

Проведенный анализ позволяет сделать следующие выводы:

1. В связной речи гласные подвержены эллипсису.
2. Сопоставление идеального и реального распределения фонем в материале показывает, что в спонтанной речи, также как и при чтении текста больше всего «чужими» аллофонами заменяются фонема /i/. Процент реализаций «своими» аллофонами в обоих видах речи выше всего у фонемы /o/.
3. Наиболее активными в процессе замен оказываются аллофоны фонем /e/ и /u/ и /a/. Количество реализаций их аллофонов на месте других фонем иногда достигает до 15 %.
4. При чтении связного текста гласный не менее, чем в 74% случаев реализуется, аллофоном, относящимся к данной фонеме. В спонтанной речи процент реализации на месте одной фонемы аллофона другой фонемы выше для всех гласных, что связано с основными свойствами спонтанной речи, ее большим разнообразием и вариативностью.

Литература

1. Бондарко Л.В. Фонетика современного русского языка. СПб., 1998.
2. Бондарко Л.В. Спонтанная речь и организация системы языка // Бюллетень фонетического фонда русского языка. № 8. Фонетические свойства русской спонтанной речи. СПб., Бохум, 2001. С. 17-23.
3. Фонетика спонтанной речи. Л., 1988.
4. Мартине А. Принцип экономии в фонетических изменениях. М., 1960.
5. Кузнецов В.И. Вокализм связной речи. СПб, 1997. С. 209.
6. Зиндер Л.Р. Общая фонетика. Л, 1979. С. 59-63.
7. Skrelin P. Labeling Problems in Speech Corpus Creation. In: Proc. of International workshop SPECOM'2005, Patras, 2005
8. Bondarko L.V., Volskaya N.B., Tananaiko S.O., Vasilieva L.A. Phonetic properties of Russian Spontaneous Speech // Proceedings of the 15th ICPHS. Barcelona, 3-9 August 2003. P. 2973-2976.
9. Ресурс Интернет. Код доступа: <http://praat.org>

С.О. Тананайко, Л.А. Васильева

Фонологическое и фонетическое описание гласных в первообразных союзах

Факультет филологии и искусств СПбГУ,
г. Санкт-Петербург, Россия
stananaiko@mail.ru, lavasil@phonetics.pu.ru

Теоретическая концепция языка как этнокультурного феномена, согласно которой звуковая форма языка рассматривается как результат сложного взаимодействия социальных, лингвистических, психофизиологических и информационных процессов, определяющих речевую деятельность носителей языка, в настоящее время получает дальнейшее развитие благодаря возможности использования современных речевых технологий для моделирования этой деятельности и архивирования звуковых форм и звуковых единиц различных иерархических уровней языка в лингвистических базах данных. Дальнейшее развитие речевых технологий позволяет исследователям обратиться к чрезвычайно существенному и интересному направлению исследований — фонетике и фонологии спонтанной речи.

Причиной повышенного внимания лингвистов к спонтанной речи в последнее время является лишь недавно сделавшаяся очевидной ее первичность по отношению к другим типам устной речи. Так, Л.В. Бондарко указывала, что рассмотрение спонтанной речи как некоего отклонения от эталонных реализаций неправомерно, так как в ситуации речевой коммуникации именно спонтанная речь является первичной [Бондарко 2001: 17]. Необходимо также иметь в виду, что в связи с быстро меняющейся социальной структурой современного российского общества получение и изучение новых данных о речевых характеристиках различных социальных групп, произношение которых может рассматриваться как норма или как узус, является чрезвычайно актуальным.

Поскольку именно спонтанная речь является источником как узуса, так и нормы, то она и стала предметом настоящего исследования, осуществляемого в рамках научно-исследовательской работы кафедры фонетики СПбГУ по проекту РФФИ 07-06-00139. Подробно методика данного исследования описаны в материалах Фонетического Конгресса в Барселоне в 2003 году. [Bondarko 2003: 2973].

В статье рассматривается один из дискуссионных вопросов русской фонетики, а именно, фонемная принадлежность и качество гласного в союзах *но, что*. Традиционно слова такого типа (первообразные союзы) рассматриваются как клитики и, следовательно, должны реализоваться по правилам безударного вокализма, но, с другой стороны, в соответствии с современной произносительной нормой слова *но, что* должны произноситься с [о]-образным гласным. Поэтому задачей данного исследования было изучение темпоральных, и перцептивных параметров гласных в этих словах, чтобы на основании полученных данных сделать вывод о преобладающих в современной нормативной речи тенденциях реализации этих гласных.

Исследование было проведено на материале русской спонтанной речи 5 принадлежащих к различным возрастным группам (от 18 до 55 лет) дикторов-носителей русского нормативного произношения. Затем на основе имеющейся спонтанной речи для всех дикторов были составлены тексты, прочитанные теми же дикторами. В спонтанной речи и чтении каждого диктора для дальнейшего анализа были выделены идентичные по словесному составу фрагменты длительностью не менее 10 минут для каждого диктора.

Была проведена полная обработка записи их речи. Вначале тексты подвергались детальному слуховому анализу. В результате была получена акустическая транскрипция, при которой транскрибировался каждый отдельный звук без учета фонетического контекста. Идеальная транскрипция создавалась в соответствии с принципами Щербовской фонологической школы.

На базе пакета программ для обработки звуковых сигналов PRAAT был разработаны скрипты, позволяющие получать различные характеристики речевого сигнала. Для каждого звука были получены 1) представление данного аллофона в 2 видах транскрипции (реальной и идеальной); 2) длительность звука.

Всего в обработанном материале встретилось 25 реализаций союза *но* и 43 реализации союза *что* в каждом типе речи. При этом не рассматривалось союзное слово *что* и *что* в составе сложных союзов. Из 86 случаев реализации союза *что* (по 43 в спонтанной речи и чтении) в 3 случаях гласный полностью выпал, и эти 3 случая были исключены из рассмотрения.

1. Темпоральные характеристики гласных

При анализе длительности гласного выяснилось, что реализации обоих союзов можно разделить на 3 группы по длительности гласного, поскольку русское ударение в первую очередь количественное и именно признак длительности является ведущим при определении ударности-безударности гласного.

Данная разбивка материала на 3 группы сделана с учетом результатов проведенного на кафедре фонетики СПбГУ всестороннего детального исследования всего корпуса спонтанной речи.

Таблица 1. Количество и средняя длительность трех групп реализаций союзов *но* и *что* в спонтанной речи и чтении

	группы	спонтанная речь		чтение	
		количество	сред. длит., мс	количество	сред. длит., мс
но	1	6	33	11	42
	2	11	66	13	67
	3	8	262	1	109
что	1	16	33	32	39
	2	13	68	6	67
	3	11	180	4	229

По длительности гласного изучаемые примеры из спонтанной речи классифицируются следующим образом (см. табл. 1):

1) Малая длительность гласного, не позволяющая уверенно опознать качество гласного. Средняя длительность гласного в этих примерах составляла 33 мс и для *но*, и для *что*.

2) Длительность гласного, позволяющая опознать его как гласный полного образования. Средняя длительность гласного в этой группе составляла 66,4 мс для *но*, 68 мс для *что*. Можно сделать вывод, что в примерах, попавших во вторую группу, реализуется ударный гласный, так как это очень близко к значениям, полученным О.Б. Болотовой. По данным О.Б. Болотовой средняя длительность ударного [о] в спонтанной речи составляет 67 мс, [Болотова 2005: 118].

3) Большая длительность гласного, связанная с хезитационной затяжкой. Длительность гласных в этой группе варьирует от 116 мс до 430 мс, составляя в среднем 262 мс для *но*, 180 мс для *что*. Большое количество примеров, попавших в эту группу, связано со спецификой спонтанной речи, для которой хезитация чрезвычайно характерна, в частности, после союзов [Степанова 2006: 99]. Кроме того, надо учитывать, что союзы *но* и *что* появляются обычно на границе простых предложений в пределах сложного (с точки зрения синтаксиса), и на границе синтагм (с точки зрения фонетического просодического членения), то есть в позициях, где хезитация наиболее частотна.

В чтении были обработаны слова из фрагментов, идентичных по словесному составу фрагментам спонтанной речи.

Примеры из чтения также можно разделить по длительности на аналогичные 3 группы:

1) Малая длительность гласного, не позволяющая уверенно опознать качество гласного. Средняя длительность гласного в этих примерах составляла 42 мс для *но*, и 39 мс для *что*.

2) Длительность гласного, позволяющая опознать его как гласный полного образования. Средняя длительность гласного в этой группе составляла 66,8 мс для *но*, 67,3 мс для *что*. Можно сделать вывод, что в примерах, попавших во вторую группу, реализуется ударный гласный, так как это очень близко к значениям, полученным О.Б. Болотовой. По данным О.Б. Болотовой средняя длительность ударного [о] в чтении — 67,2 мс [Болотова 2005: 118].

3) Большая длительность гласного, связанная с хезитационной затяжкой. Длительность гласных в этой группе варьирует от 109 мс до 299 мс, составляя в среднем 229 мс для *что*. Для *но* встретился всего один такой случай. Естественно, в чтении значительно меньше хезитаций, чем в спонтанной речи.

Интересно рассмотреть количественное распределение примеров *но* и *что* по трем группам. В спонтанной речи для слова *но* наибольшее количество примеров попадает во вторую группу, где гласный реализуется как гласный полного образования. Довольно представительна также и группа с хезитациями (третья), наименьшей оказывается группа с краткими гласными. Для слова *что* в спонтанной речи наиболее представительной оказалась первая группа — с краткими гласными, вторая группа несколько меньше, а третья группа (хезитационная) немного меньше второй.

В чтении для слова *но* преобладающей оказалась вторая группа, первая группа несколько меньше, а в третьей группе всего один пример. Для слова *что* в чтении абсолютно преобладает первая группа (32 случая), вторая и третья группы намного меньше (6 и 4 случая соответственно).

Можно сделать вывод, что наиболее предпочтительной реализацией слова *но* и в спонтанной речи, и в чтении была реализация с гласным полного образования, при этом в чтении достаточно часто встречались и краткие гласные. Для слова *что* и в спонтанной речи, и в чтении преобладает реализация с кратким гласным, причем если в спонтанной речи это преобладание незначительно, то в чтении оно абсолютно. Обращает на себя внимание небольшое количество случаев реализации гласного полного образования в союзе *что* в чтении (всего 6 случаев). Таким образом, союз *но* чаще реализуется с гласным полного образования, а союз *что* — с кратким гласным. Возможно, это связано с лексическими причинами, а именно — при редуцированном произнесении гласного в слове *но* это слово может быть опознано как предлог *на* (ср. *но* это — *на* это). Кроме того, нельзя не учитывать большей частотности *что* по сравнению с *но*. Вероятно также, что полная реализация гласного в *что* будет преобладать в союзном слове в отличие от союза.

2. Анализ формантных структур

Данные о значениях первой и второй формант свидетельствуют, что изучаемые гласные по их качеству можно разделить на три группы (см. рисунок 1). В первой группе (она обозначена на рисунках цифрой 1) значение первой форманты не выше 700 Гц, а значение второй форманты не выше 900 Гц. Это лабиализованные гласные заднего ряда. Во второй группе (цифра 2 на рисунках) значения первой форманты не выше 700 Гц, а значения второй располагаются в диапазоне 1000-1500 Гц — такие значения характерны для гласных центрального ряда. В эту же группу попали и большинство кратких неопределенных гласных. И, наконец, в третьей группе (на рисунках — цифра 3) значения второй форманты превышают 1500 Гц, следовательно, это гласные переднего ряда. Важно отметить, что гласные центрального ряда в реализации союзов *но* и *что* достаточно многочисленны, особенно для слова *что*, для которого их больше, чем гласных заднего ряда. Это объясняется включением в эту группу кратких неопределенных гласных, преобладающих, как было установлено выше, в реализациях этого слова. Гласные переднего ряда также более многочисленны в реализациях слова *что*. Это подтверждает вывод предыдущего раздела о большей редуцированности гласных в слове *что* по сравнению с *но*, в котором чаще встречались гласные полного образования.

Для исчерпывающей интерпретации полученных данных обратимся к анализу акустической транскрипции.

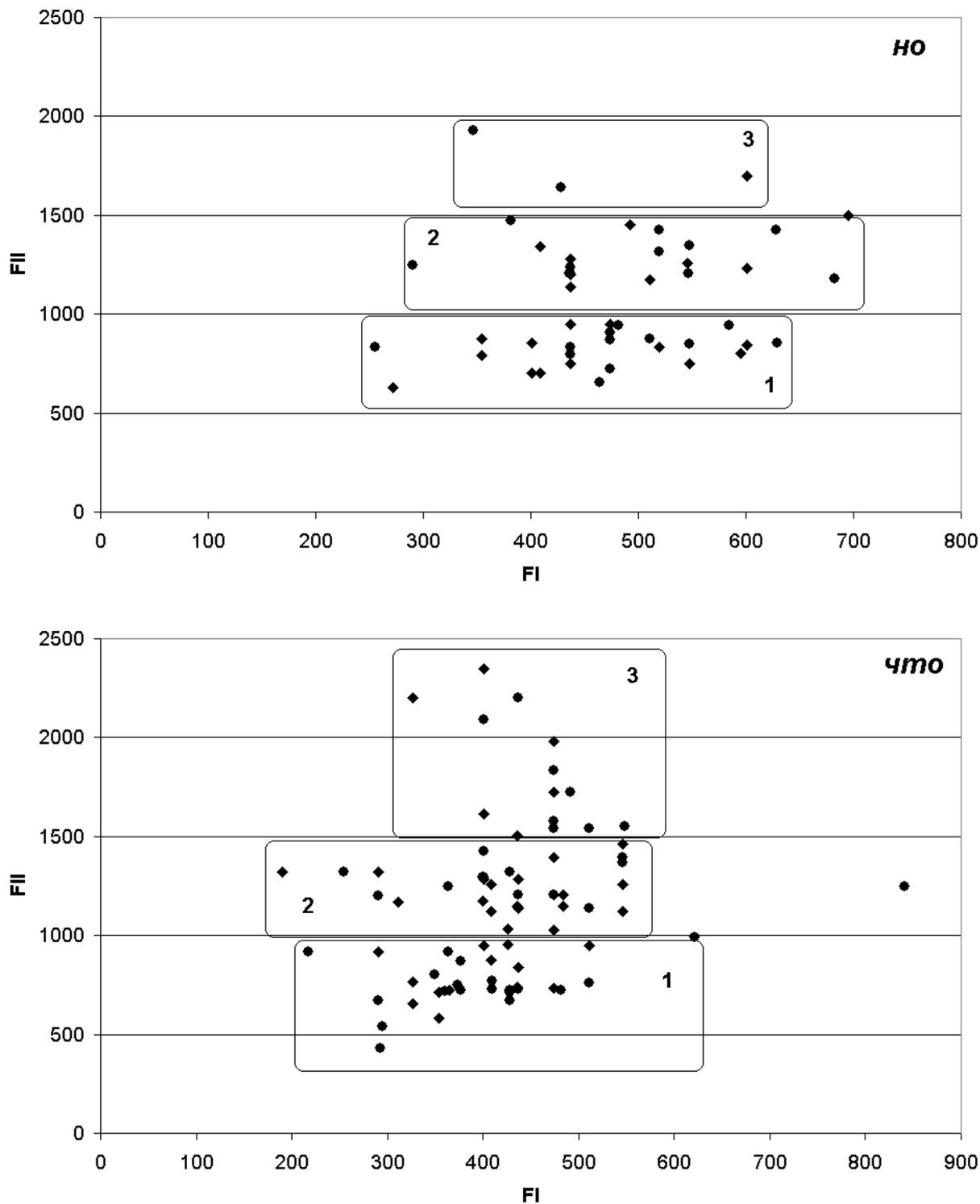


Рис.1. Значения первой (ось абсцисс) и второй (ось ординат) формант гласных в слове *но* (вверху) и слове *что* (внизу). Ромбовидный маркер \blacklozenge — данные для спонтанной речи, круглый маркер \bullet — для чтения.

3. Анализ акустической транскрипции

По данным акустической транскрипции в исследованном материале преобладают лабиализованные реализации гласных в изучаемых союзах. Эти реализации составляют 84% всех случаев *но* в спонтанной речи и 68% *но* в чтении. Для союза *что* было выявлено 76% лабиализованных реализаций в спонтанной речи и 83% в чтении. Помимо лабиализованных реализаций встречались и реализации гласных в виде аллофонов фонемы /a/, а также аллофонов фонемы /i/ (см. табл.2). Однако наиболее частотной альтернативой произнесению лабиализованных гласных была реализация неопределенного гласного центрального ряда [ə] (см. табл.2). Сопоставив эти данные с результатами анализа темпоральных характеристик, можно отметить, что абсолютное большинство гласных полного образования в изученном материале реализуется как лабиализованные, чаще всего [o]-образные гласные или лабиализованные гласные средне-верхнего подъема центрального ряда [ə]. Деллабиализованные реализации фиксируются в первую очередь для кратких гласных, хотя при этом основное количество этих гласных также реализуются как лабиализованные и чаще всего как гласные центрального ряда. Это подтверждается количественным распределением аллофонов, отмеченных на месте изучаемых гласных в акустической транскрипции (см. табл.2). Так, для слова *но* преобладающим аллофоном является [o] и в спонтанной речи, и в чтении. Это хорошо коррелирует с выводом о преобладании гласного полного образования в произношении этого союза. Для слова *что* преобладающим аллофоном оказывается лабиализованный гласный средне-верхнего подъема центрального ряда, что согласуется с выводом о преобладании кратких гласных при произношении *что*. Преобладание гласного полного образования для союза *но* может быть связано с его большей (по сравнению с союзом *что*) лексической нагруженностью: *но* часто употребляется для указания на противопоставление, а *что* часто является лишь формальным способом синтаксической связи частей предложения.

Таблица 2. Распределение аллофонов гласных в союзах *но* и *что* по данным акустической транскрипции (в % от общего количества случаев для каждого союза в каждом типе речи)

		o	ə	ɔ	ø	œ	ʊ	ə	i	a
спонт. речь	но	36	16	16	8	4	4	16	0	0
	что	25	38	0	0	3	10	17	5	2
чте- ние	но	24	16	4	0	0	24	28	0	4
	что	31	38	7	0	0	7	14	3	0

Таким образом, мы можем сделать вывод, что в союзе *но* обычно произносится [o]-образный гласный полного образования, близкий к ударному аллофону, в то время как при произношении союза *что* преобладает лабиализованный гласный центрального ряда [ə], который хотя и опознается как [o]-образный, не может считаться близким к ударному аллофону /o/, так как он слишком краткий и продвинутый вперед, что не является нормой после твердого согласного, но может рассматриваться как один из признаков безударного аллофона, которые, как известно, располагаются ближе к центру треугольника гласных, чем ударные. Кроме того, необходимо добавить, что реализация гласного в этих союзах как [o]-образного приближается к 100%, если имеется хезитационная затянутость гласного или пауза хезитации после него. Итак, можно утверждать, что в большинстве случаев и в спонтанной речи, и в чтении в союзах *но* и *что* реализуются аллофоны фонемы /o/ разной степени ударности. С точки зрения системы русского безударного вокализма это означает, что дифференциальные признаки сохраняются не только у безударных аллофонов гласных верхнего подъема, но и в некоторых случаях у безударных аллофонов гласного среднего подъема /o/, и при этом чередования ударной фонемы /o/ с безударной фонемой /a/ не происходит. А это позволяет говорить о возможности реализации безударного /o/ не только в редких заимство-

ванных словах в зияниях, но и в частотных и даже сверхчастотных словах в позиции после согласного.

Литература

1. Bondarko L.V., Volskaya N.B., Tananaiko S.O., Vasilieva L.A. Phonetic Properties of Russian Spontaneous Speech / L.V.Bondarko // Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, 3-9 August. – 2003. – P. 2973-2976
2. Бондарко Л.В. Спонтанная речь и организация системы языка / Л.В. Бондарко // Фонетические свойства русской спонтанной речи: Бюллетень Фонетического Фонда № русского языка. Под ред. Л.В. Бондарко, М. Краузе. Санкт-Петербург–Бохум, 2001. – Вып.8. – С.17-23
3. Болотова О.Б. Гласные в спонтанной речи и при чтении связного текста (экспериментально-фонетическое исследование на материале русского языка): дис... канд.фил.наук: 10.02.19. СПб, 2005. – 284 с.
4. Степанова С.Б. Соотношение синтагматического и хезитационного членения в спонтанной русской речи / С.Б. Степанова // Инновации в исследованиях русского языка, литературы и культуры: сборник докладов. // Пловдив: Университетско издателство «Паисий Хилендарски», 2006. – Том 1. – С

Б.М. Лобанов,* Л.И. Цирульник**

Автоматизация анализа просодических характеристик для экспериментальных исследований и синтеза речи по тексту

Объединённый институт проблем информатики Национальной академии наук Беларуси,
г. Минск, Беларусь

*lobanov@newman.bas-net.by, **l.tsirulnik@newman.bas-net.by

1. Введение

Просодические характеристики речи варьируются в зависимости от типа произносимого текста, стиля речи, а также персональных характеристик диктора — как постоянных, так и ситуативных. Исследованиям просодических характеристик речи посвящены работы [1], [2]. Отличительной особенностью этих исследований является то, что в них показаны явления, которые имеют место в естественной речи, однако в них не предлагается какого-либо механизма для использования результатов исследований при синтезе речи по тексту.

Тем не менее, существует достаточно большое количество просодических моделей, используемых в системах синтеза речи и способных адекватно отражать важные для слухового восприятия просодические характеристики естественной речи. К основным просодическим моделям относятся:

- автосегментная метрическая модель [3], представляющая интонацию в виде линейной последовательности уровней тона;
- IPO-модель [4], которая представляет интонацию как дискретную последовательность движений основного тона;
- суперпозиционная модель [5], которая основана на физиологической модели речеобразования [5], объединенной с иерархической теорией просодической фонологии [6], и представляет интонацию как последовательность событий с перекрывающимися областями;
- непрерывная параметрическая модель [7], которая интерпретирует частоту основного тона в акустической области, описывая интонацию в терминах движения F_0 во времени;
- модель портретов акцентных единиц (ПАЕ) [8], согласно которой каждое предложение состоит из последовательности синтагм. Каждая синтагма, в свою очередь, состоит из одной или более акцентных единиц. Акцентная единица (АЕ) является минимальной просодической единицей и состоит из одного или более слов, имеющих лишь один полноударный гласный. Интонационно значимыми элементами АЕ являются: ядро (полноударный гласный), предъядро (все фонемы, предшествующие полноударному гласному), и заядро (все фонемы, следующие за полноударным гласным). Основное предположение ПАЕ-модели в том, что топологические свойства просодических параметров не зависят от конкретного фонетического контекста и количества слогов в предъядре и заядре для конкретного типа интонации. Таким образом, просодические характеристики могут задаваться «портретами» акцентных единиц, которые указывают нормированные значения F_0 на участках предъядра, ядра и заядра.

ПАЕ-модель описывает интонацию в терминах движения F_0 во времени, но характер движения не ограничивается фиксированными параметрами, а может быть выбран исходя из конкретной реализации определённого типа интонации с учётом стиля речи и персональных особенностей диктора. Кроме того, ПАЕ-модель распространяется на описание движений амплитуды и длительности звуков.

В данной работе описываются две компьютерные системы, позволяющие автоматически анализировать естественную речь и создавать мелодические, энергетические и ритмические портреты акцентных единиц в рамках ПАЕ-модели, которые эффективно используются при синтезе речи по тексту.

2. Общая схема подготовки и обработки данных для создания портретов акцентных единиц

Схема подготовки и обработки данных показана на рисунке 1. На начальном этапе опытный фонетист анализирует фонограмму, разбивает её на синтагмы (под синтагмой понимается самостоятельная в интонационном смысле часть фразы или вся фраза) и создаёт стенограмму записей. Решение о конце синтагмы принимается на основе ряда признаков, таких как: присутствие дыхательной паузы, комплексная реализация одного из возможных интонационных типов синтагмы, наличие определённой динамической структуры (контура силы звука) и определённой ритмической структуры (контура длительности звуков). При создании стенограммы учитывается то, что некоторые слова могут быть неверно произнесены диктором, а ударения могут не соответствовать каноническим ударениям. Фонетист стенографирует слова именно так, как они произнесены диктором, и помечает позиции ударения в случаях, когда они не являются каноническими. Результатом ручной обработки фонограммы являются списки речевых и соответствующих текстовых синтагм, которые подаются на вход компьютерной системы «ИнтоКлонатор».

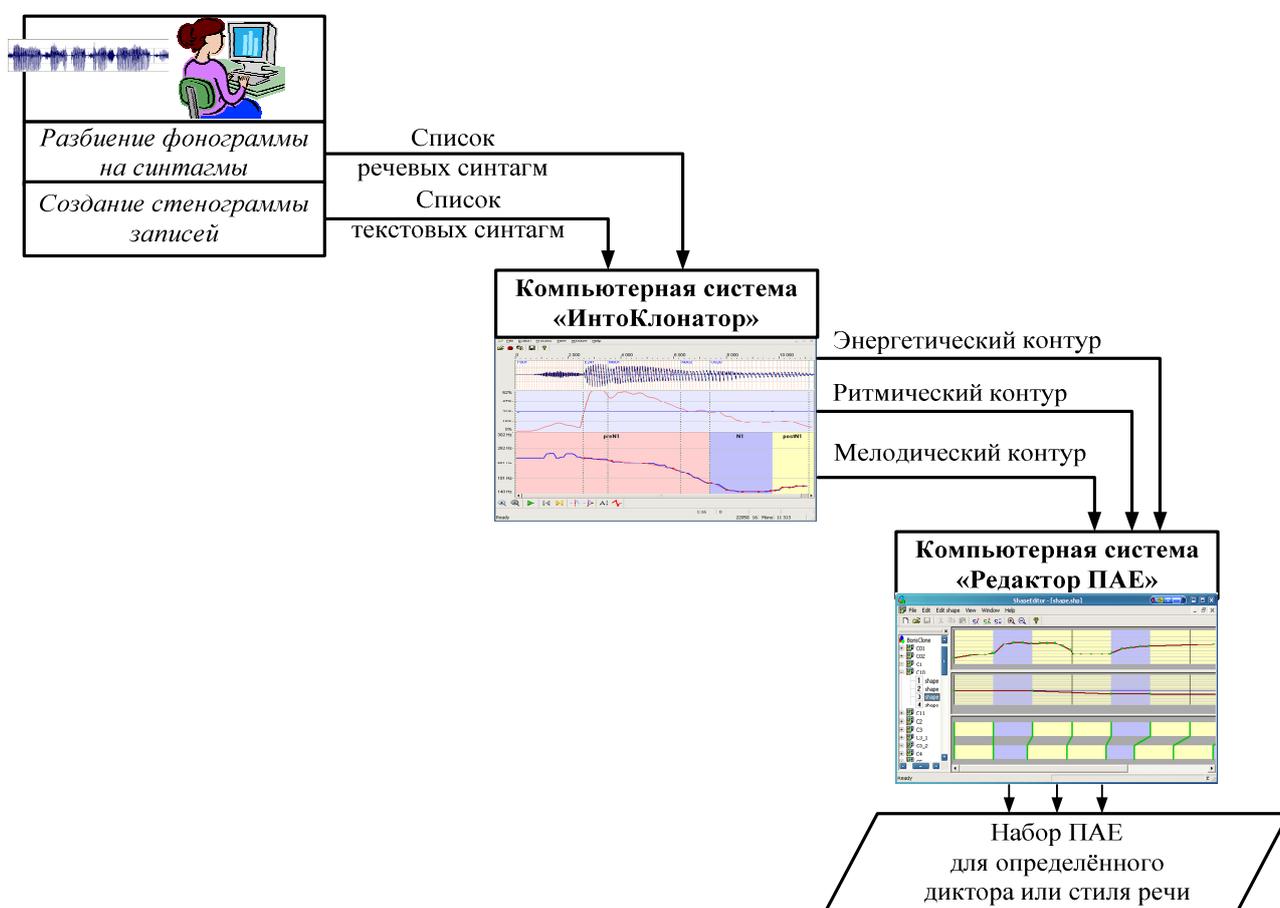


Рис. 1. Схема подготовки и автоматической обработки данных для создания ПАЕ.

В системе «ИнтоКлонатор» [9] для каждой синтагмы автоматически выполняются следующие действия:

- синтез речевого сигнала на основе входной стенограммы записей; при этом в синтезированном сигнале помечаются границы аллофонов, АЕ и элементов АЕ (ЭАЕ): предъядра, ядра и заядра;
- анализ акустических признаков естественного и синтезированного сигналов, их сопоставление методом динамического программирования и перенос маркеров границ аллофонов, АЕ и ЭАЕ с синтезированного на естественный речевой сигнал;
- разметка естественного сигнала на периоды основного тона (питчи);

- вычисление энергетического контура на основе разметки синтагмы естественной речи на аллофоны, АЕ и ЭАЕ с помощью алгоритмов обработки амплитуды аллофонов;
- вычисление ритмического контура на основе разметки синтагмы естественной речи на аллофоны, АЕ и ЭАЕ с помощью алгоритмов обработки длительности аллофонов;
- вычисление мелодического контура на основе разметки синтагмы естественной речи на аллофоны, АЕ и ЭАЕ, а также разметки на питчи с помощью алгоритмов обработки частоты основного тона сигнала.

Результаты работы компьютерной системы «ИнтоКлонатор» могут использоваться для экспериментальных исследований просодических характеристик различных стилей речи и (или) различных дикторов.

Кроме того, сформированные энергетический, ритмический и мелодический контуры являются входными данными компьютерной системы «Редактор ПАЕ», результатом работы которой является набор просодических портретов акцентных единиц для определённого диктора или стиля речи. В системе «Редактор ПАЕ» автоматически выполняются следующие действия:

- на основе мелодического контура создаётся мелодический портрет синтагмы определённого типа, при этом значения частоты основного тона на вокализованных участках речевого сигнала интерполируются на невокализованные участки, и осуществляется нормализация портрета по длительности и частоте основного тона [10];
- на основе энергетического контура создаётся энергетический портрет синтагмы определённого типа, при этом принимаются во внимание значения амплитуды сигнала на участках ядра каждой АЕ, при нормировании учитываются «стандартные» соотношения амплитуд ударных гласных, полученные на основе анализа естественной речи;
- на основе ритмического контура создаётся ритмический портрет синтагмы определённого типа, при этом анализируются относительные длительности участков ядра каждой АЕ.

Необходимо отметить, что созданные просодические портреты могут использоваться затем системой синтеза речи по тексту «Мультифон» [11] для генерации мелодических, энергетических и ритмических характеристик синтезированных речевых синтагм.

Дополнительным достоинством компьютерной системы «Редактор ПАЕ» является возможность вручную изменить созданные просодические портреты и, прослушав синтезированную речь, определить, насколько внесённые изменения важны для слухового восприятия просодики.

3. Пользовательский интерфейс компьютерной системы «ИнтоКлонатор»

Пользовательский интерфейс системы «ИнтоКлонатор» (рисунок 2) включает следующие блоки:

- окно отображения осциллограммы речевого сигнала (РС);
- окно отображения графика амплитуды (A) сигнала;
- окно отображения графика ЧОТ (F_0) сигнала;
- диалоговые окна настроек параметров системы.

На осциллограмме РС указываются границы периодов основного тона и аллофонов, а также имена аллофонов. В системе реализовано масштабирование отображения и прослушивание выделенного фрагмента РС. На графиках отображаются границы АЕ синтагмы и предъядра, ядра и заядра каждой АЕ, а также вычисленные динамический и мелодический контуры.

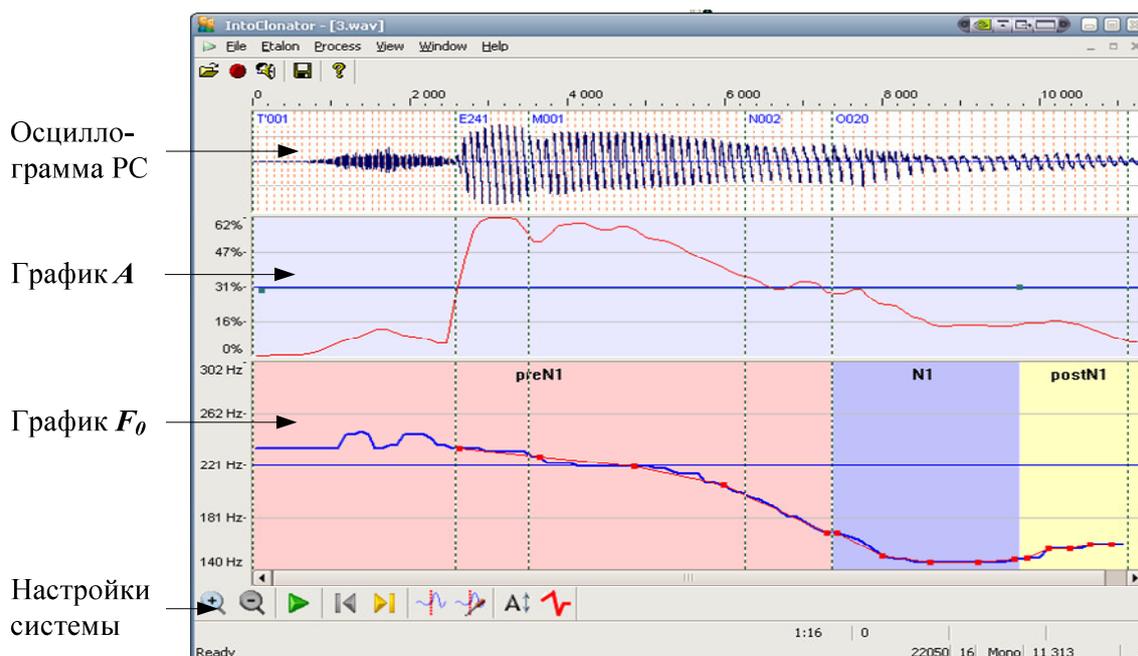


Рис. 2. Общий вид пользовательского интерфейса системы «ИнтоКлонатор».

Настройки параметров системы реализованы в следующих диалоговых окнах:

- диалог настроек параметров вычисления ЧОТ;
- диалог настроек параметров сегментации РС,
- диалог установки диапазона амплитуды и ЧОТ.

Настройки блока вычисления ЧОТ позволяют устанавливать параметры вычисления спектральных характеристик и параметры определения вокализованных участков сигнала.

Настройки блока сегментации позволяют устанавливать параметры сопоставления естественного и синтезированного РС.

Настройки диапазона амплитуды и ЧОТ позволяют указывать значения A_{min} , A_{max} и $F_{0 min}$, $F_{0 max}$, которые варьируются в зависимости от диктора и стиля речи и должны быть определены заранее для набора речевых синтагм, обрабатываемых системой.

4. Пользовательский интерфейс компьютерной системы «Редактор ПАЕ»

Пользовательский интерфейс системы «Редактор ПАЕ» (рисунок 3) включает следующие блоки:

- окно выбора просодического стиля и просодического портрета синтагмы определённого типа внутри выбранного стиля;
- окно просмотра/редактирования мелодического портрета (F_0);
- окно просмотра/редактирования энергетического портрета (A);
- окно просмотра/редактирования ритмического портрета (T).

На каждом из просодических портретов указаны границы АЕ, а также границы предъядра, ядра и заядра каждой АЕ. У пользователя есть возможность менять положение точек на графиках мелодического и энергетического портретов, а также добавлять и удалять любое количество точек. Кроме того, пользователь может изменять относительную длительность любого элемента АЕ на графике ритмического портрета.

Система позволяет создавать новые просодические стили, а также менять количество и типы просодических портретов внутри выбранного стиля.

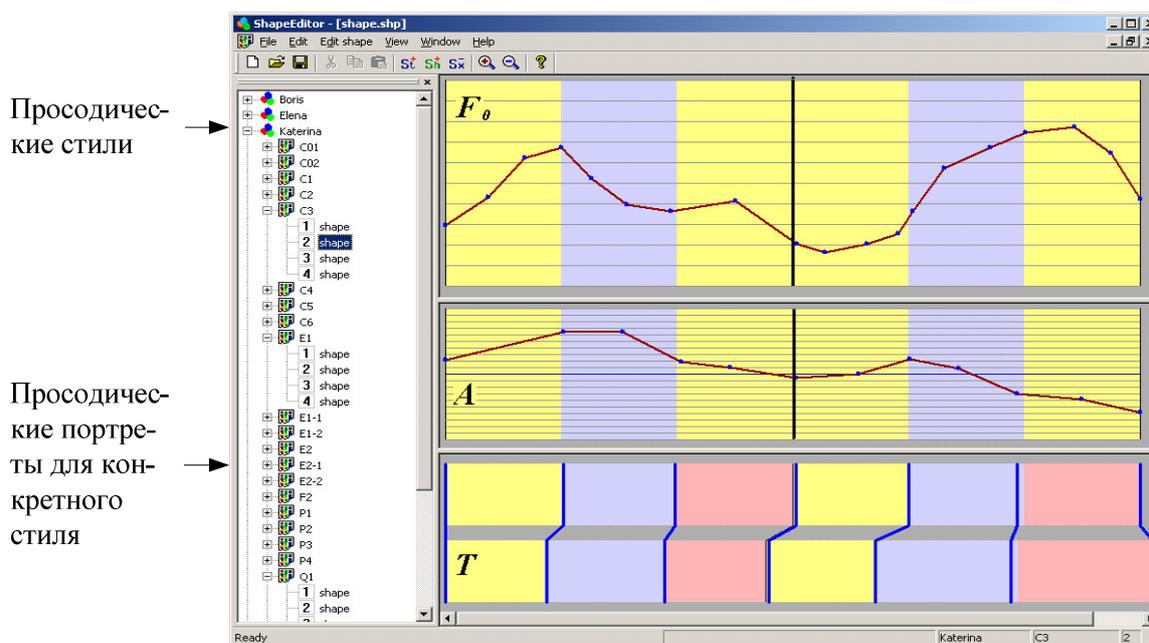


Рис.3. Общий вид пользовательского интерфейса системы «Редактор ПАЕ»

5. Использование компьютерных систем «ИнтоКлонатор» и «Редактор ПАЕ» для анализа просодических характеристик различных стилей речи

Примеры формирования просодических контуров и просодических портретов для стилей речи «чтение художественного текста» и «репортаж» представлены на рисунках 4, 5.

Рисунок 4 иллюстрирует просодические характеристики синтагмы «и радостно захлопала в ладошки», произнесённой диктором-женщиной. Данная синтагма содержит две акцентные единицы: «и= ра+достно / захло=пала в ладо+шки» (здесь и далее знаком «/»отмечены границы АЕ, знаком «=» после ударного гласного — словесное ударение, знаком «+» после ударного гласного — синтагматическое ударение). На основании анализа фонограмм речи данного диктора установлен следующий диапазон изменения F_0 : $F_{0\min} = 125$ Гц, $F_{0\max} = 300$ Гц.

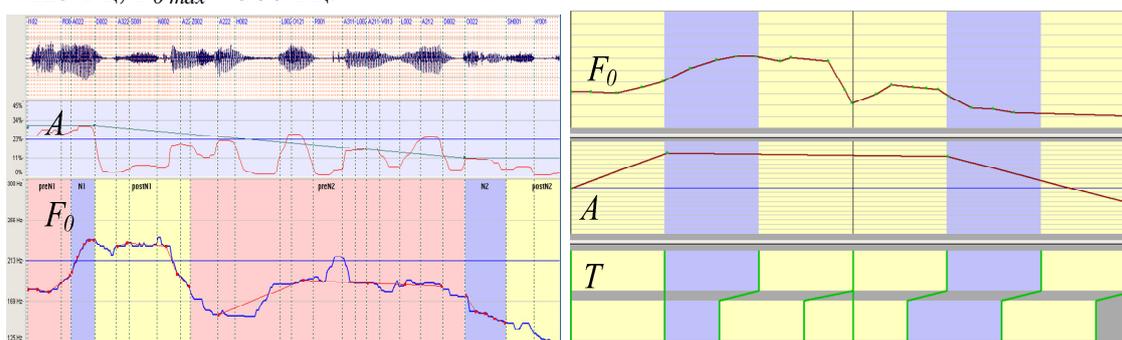


Рис.4. Просодические контуры и просодические портреты двухакцентной синтагмы «и= ра+достно / захло=пала в ладо+шки»

На рисунке 5 показаны просодические характеристики синтагмы «из-за компьютерного сбоя в системе идентификации», произнесённой профессиональным корреспондентом НТВ (мужской голос). Данная речевая запись является фрагментом базы данных «Интонация русских информационных текстов». Синтагма состоит из двух акцентных единиц: «из-за компью=терного сбоя / в систе=ме идентифика+ции»; диапазон изменения F_0 : $F_{0\min} = 80$ Гц, $F_{0\max} = 240$ Гц.

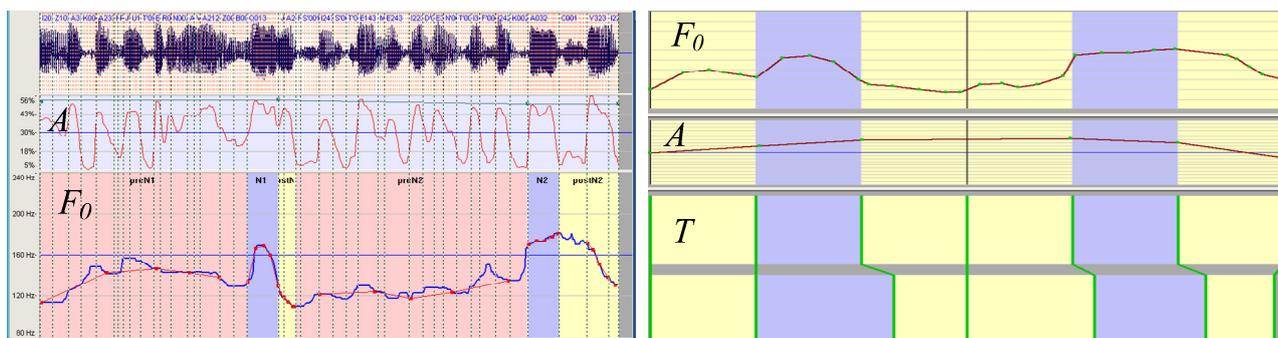


Рис.5. Просодические контуры и просодические портреты двухакцентной синтагмы
«из-за компью=терного сбо+я / в систе=ме идентифика+ции»

5. Заключение

Предлагаемая в работе технология подготовки и обработки речевых данных позволяет существенно сократить время и усилия, необходимые для анализа просодических характеристик естественной речи.

Опыт использования компьютерных систем «ИнтоКлонатор» и «Редактор ПАЕ» для анализа и синтеза просодических характеристик делового и художественного стилей речи подтвердил их высокую эффективность. Предварительные исследования показали, что просодическая модель ПАЕ способна также адекватно отражать просодические явления разговорной речи.

Литература

1. Светозарова, Н.Д. Акцентно-ритмические интонации в русской спонтанной речи. Проблемы фонетики: сб. статей. М.: Прометей, 1993, с. 189-198.
2. Кодзасов, С.В. Размещение тональных акцентов в русском слове. Язык и речь: проблемы и решения: сб. науч. трудов. М.: МАКС пресс, 2004, с. 62-76
3. Silverman K. et al. TOBI: a standard for labelling english prosody. Spoken Language Processing. ICSLP'92, Alberta, 1992, p. 867–870.
4. Hart, J., Collier, R., Cohen, A. A Perceptual study of intonation: an experimental phonetic approach to speech melody. Cambridge: Cambridge University Press, 1990, 227 p.
5. Fujisaki, H. Dynamic characteristics of voice fundamental frequency in speech and singing. The Production of Speech. NY: Springer-Verlag, 1983, ch. 3, p. 39–47.
6. Nespor, M., Vogel, I. Prosodic phonology. Dordrecht: Foris Publications, 1986, 180 p.
7. Taylor, P. Analysis and synthesis of intonation using the tilt model. J. Acoust. Soc. America, 2000, vol. 107, No 3, p. 1697–1714.
8. Lobanov, B. The phonemophon text-to-speech system. ICPHS'87, Tallin, 1987, vol.1, p. 120–124.
9. Лобанов, Б.М., Цирульник, Л.И., Сизонов, О.Г. «ИнтоКлонатор» — компьютерная система клонирования просодических характеристик речи. «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2008, с. 330-338.
10. Цирульник, Л.И., Жадинец, Д.В., Лобанов, Б.М., Сизонов, О.Г. Алгоритмы синтеза просодических характеристик речи по тексту в системе "Мультифон". «Компьютерная лингвистика и интеллектуальные технологии». М.: Издательский центр РГГУ, 2007, с. 550-558.
11. Лобанов, Б.М. «МультиФон» — система персонализированного синтеза речи по тексту на славянских языках. Лингвистическая полифония. М.: Языки славянских культур, 200, с. 849-866.
12. Кодзасов, С.В., Архипов А.В., Захаров Л.М., Кривнова О.Ф. База данных «Интонация русских информационных текстов». «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2008, с. 206-209.

А.А. Конев, Р.В. Мещеряков***

Оценка вокализованных участков речевого сигнала

*Томский государственный университет систем управления и радиоэлектроники,
г. Томск, Россия,*

**kaa1@keva.tusur.ru, **mrv@keva.tusur.ru*

1. Введение

При построении систем распознавания слитной речи с неограниченным словарем важное значение имеет качественная предварительная обработка сигнала. Очевидно, что точность выделения таких параметров, как частота основного тона, интенсивность гармоник, динамика их изменения и т.д. непосредственно зависит от этого этапа обработки речевого сигнала. Использование модели обработки речевого сигнала на периферической части слуховой системы человека является одним из наиболее перспективных методов предварительной обработки.

Рассматриваемая модель включает в себя фильтрацию, одновременную маскировку сигнала и получение его параметров в координатах интенсивность-время-частота. Фильтрация осуществляется при помощи набора фильтров, учитывающих известные из психоакустики характеристики слуховой системы человека [1]. При этом для обработки используется окно Гаусса.

В работе представлены результаты одновременной маскировки речевого сигнала, результаты сегментации на вокализованные и невокализованные участки, динамика изменения частоты основного тона. В качестве примера использовалась фраза «Эти жирные сазаны ушли под палубу», произнесенная женским и мужским голосами. Сигналы записывались с частотой дискретизации 8 кГц, частотный диапазон анализа — 50-2200 Гц.

2. Одновременная маскировка

Предварительная фильтрация по частотным каналам, количество которых в примере равно 128, позволяет получить следующие параметры речевого сигнала: массив интенсивностей и массив частот для каждого дискретного отсчета времени.

Следующий этап обработки речевого сигнала — одновременная маскировка. Она позволяет выделить составляющие сигнала, которые воспринимаются слуховой системой человека [2]. Таким образом, после одновременной маскировки появляется ещё одно представление сигнала, где, кроме координат время-частота, существует координата, принимающая бинарные значения: 0 — "не воспринимается слуховой системой", 1 — "воспринимается слуховой системой".

На рисунках 1 и 2 приведены результаты одновременной маскировки для речевого сигнала примера. Достаточно четко прослеживается периодическая структура на вокализованных участках входного сигнала. Черным цветом представлены составляющие речевого сигнала, воспринимаемые слуховой системой (значения 1 после одновременной маскировки). По оси абсцисс отложено время в секундах, по оси ординат — номера частотных каналов фильтрации. При этом канал с номером 127 соответствует частоте 50 Гц, а канал с номером 0 — частоте 2200 Гц.

Анализ полученных результатов одновременной маскировки позволяет определять номер частотного канала фильтрации, соответствующий основному тону сигнала [3]. При этом создаются бинарные шаблоны, включающие срез сигнала после одновременной маскировки. Каждый шаблон включает две гармоники и создается для своего номера частотного канала фильтрации. На каждом временном отсчёте после свертки набора шаблонов с массивом значений, полученных после одновременной маскировки, вычисляется мера различия каждого шаблона и этого массива. Далее выбирается шаблон с наименьшим значением меры разли-

чия. Если полученное значение меры различия не превышает заданный порог вокализации, то номер канала, для которого создавался выявленный шаблон, соответствует основному тону речевого сигнала. Таким образом, временные отсчеты, на которых порог вокализации не был превышен, признаются вокализованными.

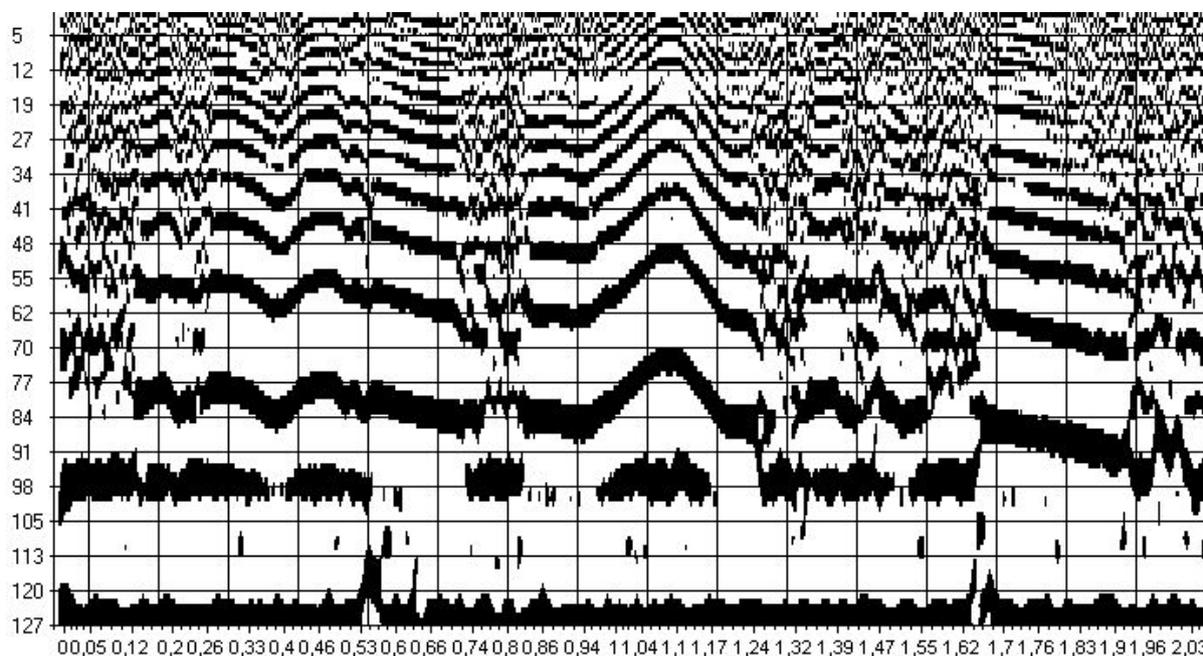


Рис. 1. Результат одновременной маскировки (женский голос).

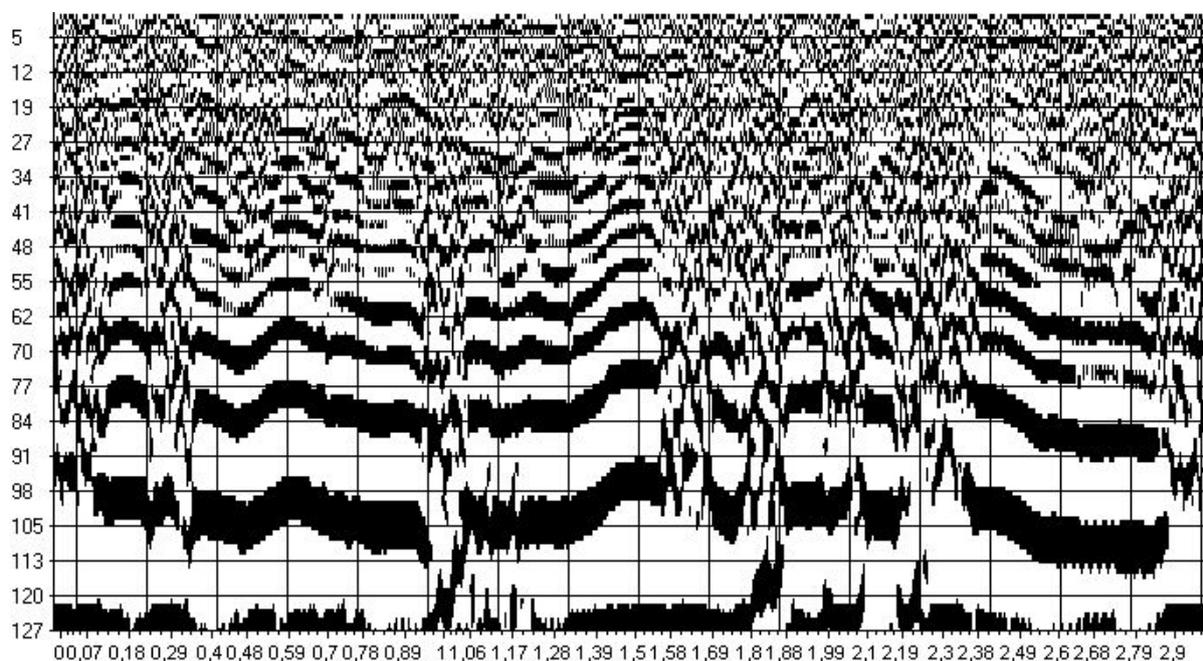


Рис. 2. Результат одновременной маскировки (мужской голос).

3. Сегментация на вокализованные и невокализованные участки

Для проведения сегментации необходимо найти непрерывные отрезки сигнала, состоящие из дискретных отсчетов с одинаковым признаком вокализации. Введены ограничения на минимальную длительность вокализованных и невокализованных участков. Из психоакустики известно, что для восприятия высоты звуков речи человеку требуется как минимум 20-30 мс [1], что и послужило выбором значений ограничений. Для вокализованных сегментов минимальная длительность была принята за 30 мс. При обнаружении участка, состоящего только из «вокализованных» отсчетов, его длительность сравнивается с минимально возможной.

Если минимальная длительность оказывается больше, то обнаруженный участок признаётся невокализованным. Аналогичный алгоритм используется и для обнаруженных невокализованных участков. Используемый алгоритм сегментации подробно описан в [4].

В таблицах 1 и 2 приведены значения на определенные моменты времени, полученные при автоматической сегментации границы между вокализованными и невокализованными участками, а также их сравнение с эталонной ручной сегментацией. В результате было установлено, что основные ошибки возникают при определении границы заключительного сегмента.

Таблица 1. Сегментация на вокализованные и невокализованные участки (женский голос).

Транскрипция	Ручная сегментация, с	Автоматическая сегментация, с	Разность, с
«пауза»	0	0	
Э	0,14	0,14	0
т'	0,21	0,20	0,01
ижЫрный'э	0,26	0,25	0,01
с	0,73	0,72	0,01
азАныу	0,83	0,83	0
ш	1,24	1,24	0
л'и	1,33	1,33	0
п	1,40	1,39	0,01
а	1,46	1,47	0,01
тп	1,53	1,52	0,01
Алубу	1,65	1,65	0
«пауза»	1,99	1,90	0,09

Таблица 2. Сегментация на вокализованные и невокализованные участки (мужской голос).

Транскрипция	Ручная сегментация, с	Автоматическая сегментация, с	Разность, с
«пауза»	0	0	
Э	0,17	0,18	0,01
т'	0,28	0,28	0
ижЫрный'э	0,40	0,40	0
с	0,97	0,98	0,01
азанЫ	1,10	1,1	0
«пауза»	1,60	1,59	0,01
у	1,72	1,72	0
ш	1,78	1,78	0
л'и	1,91	1,91	0
п	2,00	2,00	0
а	2,11	2,11	0
тп	2,18	2,18	0
Алубу	2,39	2,39	0
«пауза»	2,86	2,84	0,02

4. Определение частоты основного тона

На сегментах, определенных как вокализованные, производится определение частоты основного тона речевого сигнала. Для этого в каждый дискретный момент времени вычисляется мгновенная частота сигнала на канале фильтрации, определенном как канал, соответствующий основному тону. Эксперименты на синтезированных периодических сигналах показали, что погрешность вычисления частоты основного тона составляет не более 0,6 % [4].

На рисунках 3 и 4 приведена динамика изменения частоты основного тона для рассматриваемых в примере сигналов. По оси абсцисс отложено время в секундах, по оси ординат — частота в Герцах. Оценка значений частоты основного тона позволяет формировать основу для гармонического анализа вокализованных сегментов.

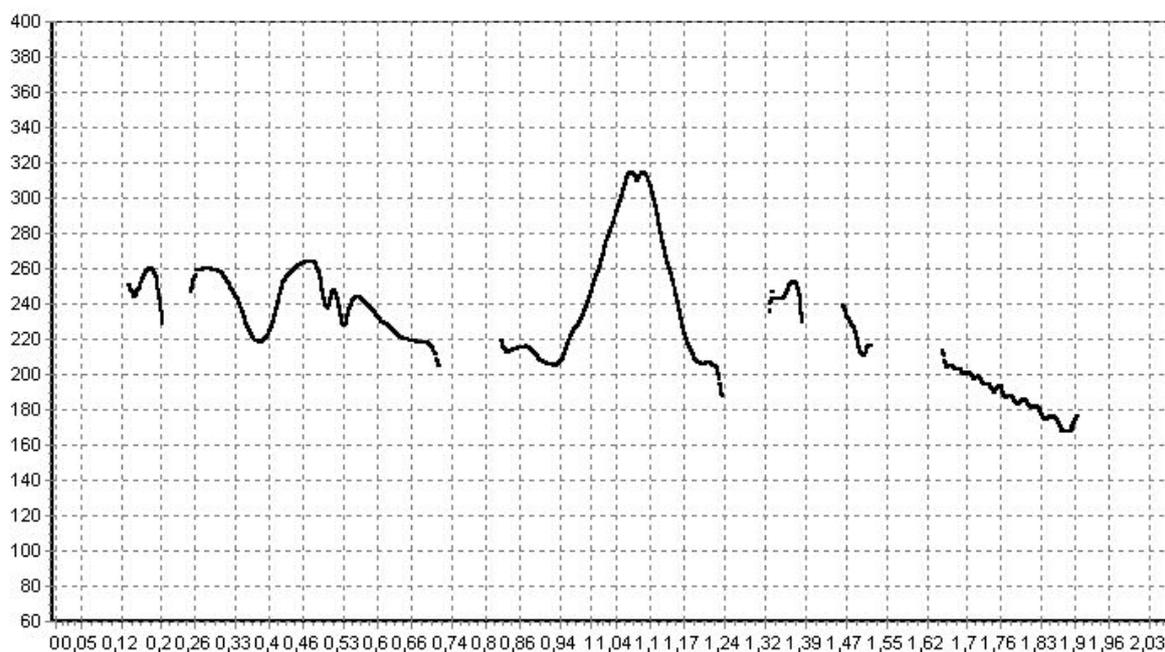


Рисунок 3. Динамика изменения частоты основного тона (женский голос).

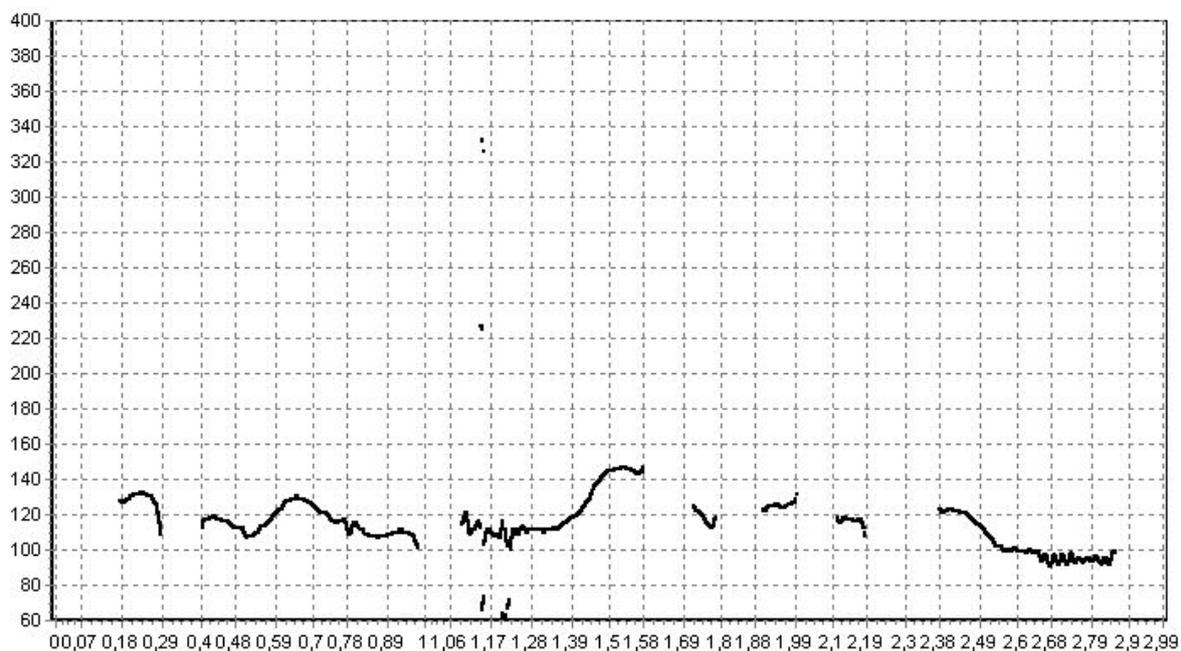


Рисунок 4. Динамика изменения частоты основного тона (мужской голос).

5. Определение характеристик частот гармоник

После определения номера канала фильтрации, соответствующего основному тону сигнала, появляется возможность вычислить номера каналов остальных гармоник в выбранном частотном диапазоне анализа. Каждому дискретному отсчету времени, определенному как вокализованный, соответствует свой набор параметров. Он содержит массивы интенсивностей и частот для номеров каналов фильтрации, соответствующих гармоникам сигнала.

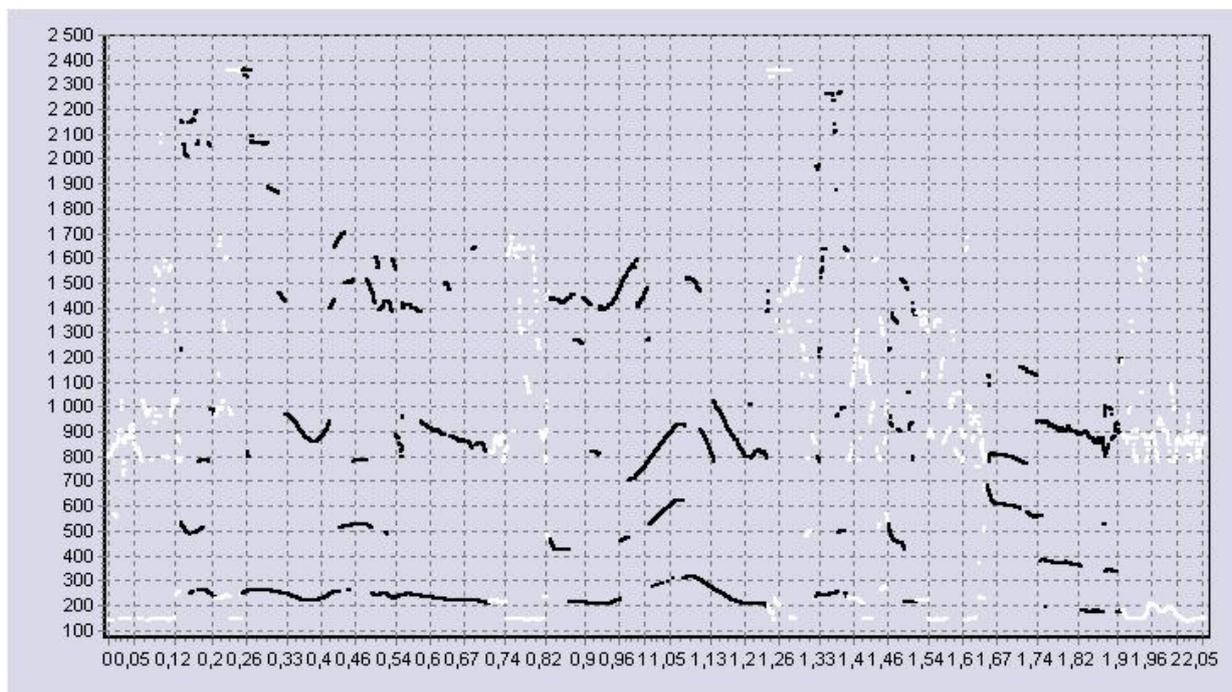


Рисунок 5. Максимумы интенсивности сигнала (женский голос).

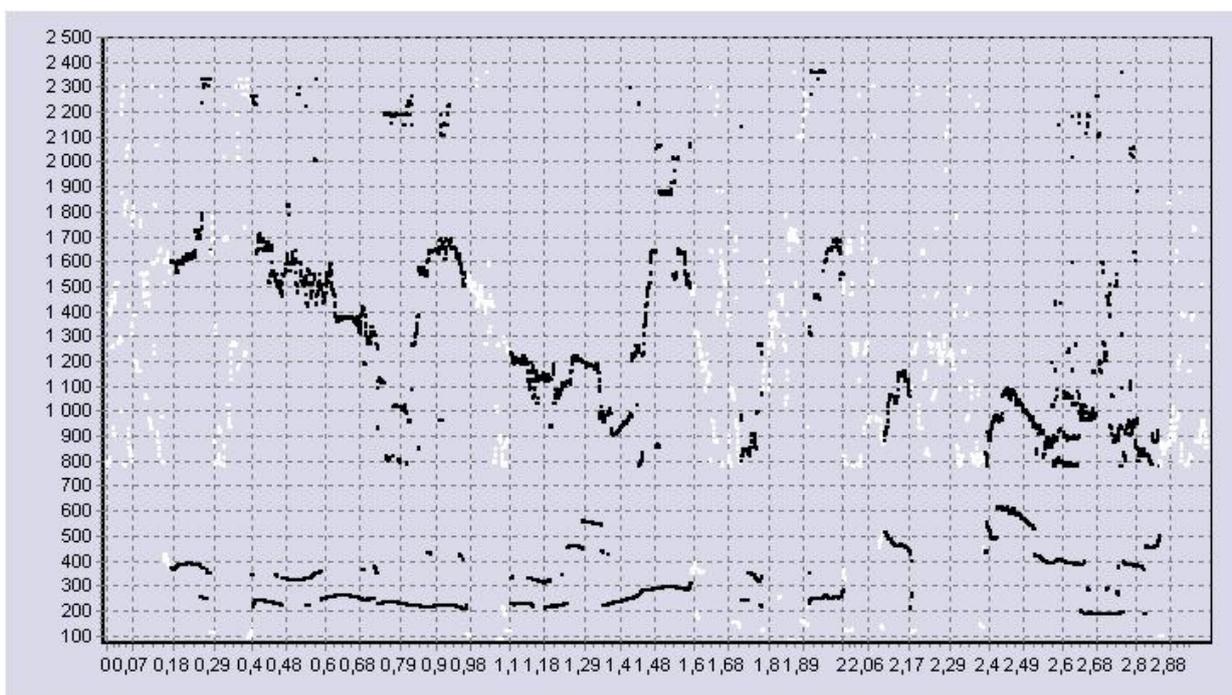


Рисунок 6. Максимумы интенсивности сигнала (мужской голос).

На рисунках 5 и 6 представлены графики определения частот речевого сигнала (изображены черным цветом), являющихся максимальными по интенсивности в двух частотных областях: от 50 до 800 Гц и от 800 до 2500 Гц. Значения частот, полученные на соответствующих гласным звукам сегментах, по своему значению близки к частотам первой и второй формант гласных, приведенным в [5].

Для невокализованных сегментов на рисунках 5 и 6 приведены определенные в указанных двух частотных областях значения частот, на которых интенсивность сигнала максимальна (изображены белым цветом).

Таким образом, выделение максимальных интенсивностей по выбранным сегментам является оценкой вокализованных участков и по сути может являться параметрическим опи-

санием гласных звуков. Для невокализованных участков значения каналов максимальной интенсивности могут быть использованы для определения энергии шумовой составляющей и ее распределения в частотной области.

6. Заключение

Предложенная последовательность обработки позволяет сегментировать речевой сигнал на два основных класса участков — вокализованные и невокализованные. Подобное разделение может являться основой для использования различных алгоритмов дальнейшей обработки и определять каждый из участков своим набором параметров. В лучшем случае сегментированные участки будут соответствовать звуком речевого сигнала. Так, для вокализованных звуков набор параметров включает интенсивность и частоту гармоник сигнала. При этом для последующей классификации целесообразно использовать динамику изменения частоты гармоник, максимальных по интенсивности в какой-либо заданной области. Погрешность определения частотных характеристик составляет менее 1%, что позволяет с достаточной степенью уверенности использовать предложенную методику для определения просодических характеристик речевого сигнала при анализе слитной речи.

Литература

1. Бекеш Г., Механические свойства уха // Бекеш Г., Розенблат В.А. // Экспериментальная психология: в 2 т. М.: Иностранная литература, 1963. – Т. 2. – С. 682-723.
2. Бондаренко В.П., Модель одновременной маскировки // Бондаренко В.П., Пономарев А.А., Рогозинская Е.А. // Интеллектуальные системы в управлении, конструировании и образовании. Томск: SST, 2004. – С. 167-174.
3. Конев А.А., Выделение вокализованных звуков в слитной речи // Конев А.А., Тихонова В.И. // Сборник трудов XVI сессии Российского акустического общества. Т. III. М.: ГЕОС, 2005. – С. 47-50.
4. Бондаренко В.П., Обработка речевых сигналов в задачах идентификации // Бондаренко В.П., Конев А.А., Мещеряков Р.В. // Изв. вузов. Физика. 2006. Т. 49, № 9. – С. 207-210.
5. Бондарко Л.В., Звуковой строй современного русского языка. // Бондарко Л.В. М.: Просвещение, 1977, 175 с.

А.В. Венцов

Спонтанная речь и ментальный лексикон

Санкт-Петербургский государственный университет,
г. Санкт-Петербург, Россия
av.ventsov@mail.ru

Ранее, описывая результаты опытов по восприятию спонтанной речи, мы предположили, что успешная идентификация редуцированных форм может быть результатом обращения к частотному словарю словоформ уже на самых ранних этапах процесса восприятия речи (Венцов и др., 2007). Отсюда с неизбежностью следует, что на первых (входных) уровнях ментального лексикона (в перцептивном словаре) его элементы должны быть описаны в терминах аллофонов (Венцов, 2007), поскольку именно с таким (акустико-фонетическим) описанием редуцированных форм приходится иметь дело системе восприятия речи на начальных стадиях обработки речевого сигнала.

Возникает вопрос, о каком аллофонном описании должна идти речь: полученном в результате транскрибирования письменного представления соответствующих элементов словаря, исходя из теоретических правил, или отражающем статистику реальной спонтанной речи. Вероятно, предпочтительнее следует второй вариант, поскольку теоретическая транскрипция никак не учитывает возможную редукцию словоформ в спонтанной речи. К сожалению, на сегодняшний день отсутствуют сколько-нибудь систематизированные сведения о частотности редуцированных форм в русской речи, а потребность в такого рода информации становится все более ощутимой, особенно применительно к исследованиям и моделированию процессов восприятия речи человеком.

Чтобы хоть немного восполнить этот пробел, я попытался проанализировать записи диалогов из телевизионной передачи «Культурная революция», сделанные на филологическом факультете СПбГУ в рамках программы исследования спонтанной русской речи (Горбова и др., 2006). Изначально меня интересовали фонетические особенности, различающие (или объединяющие) пары слов типа «знаменательное слово – частица *не* + знаменательное слово». Например, *необъятный* – *не обижай*. Однако в процессе работы мне показалось, что результаты анализа элементов этих пар могут быть интересны с точки зрения способа их возможного представления в перцептивном словаре даже вне связи друг с другом.

На рис. 1 приведены динамические спектрограммы слов *необходим(а, ы, ые, ый)* в произнесении пяти дикторов-мужчин. Очевидной особенностью этих спектрограмм является полное отсутствие сегмента, соответствующего второму предударному «о». Тем не менее, при прослушивании записей его отсутствие обнаружить невозможно (phoneme restoration effect). И объяснить это можно тем, что в современном русском языке последовательность символов *не...бх* встречается только в словах этого класса. Не следует ли из этого, что в перцептивном словаре данная группа словоформ может быть представлена и без второго предударного «о»?

Обратимся теперь к спектрограммам еще нескольких слов (рис. 2). На спектрограмме слова *неотделимы* так же, как на спектрограммах рисунка 1, отсутствует второе предударное «о». Но, как и в первом случае, это не приводит к ошибочной идентификации данного слова при произнесении в контексте. Видимо, можно говорить о том, что эти слова относятся к одному типу по способу их предполагаемого представления в словаре. Совершенно иначе выглядит спектрограмма слова *неокрашенная*: четко просматривается сегмент, соответствующий редуцированному первому предударному «о» [а]. Этому может быть очевидное объяснение — утрата этого сегмента приведет к появлению другого слова (*некрашенная*).

Наибольший интерес, как мне кажется, представляет третья спектрограмма на рис. 2: слово *необоснованно*. Второе предударное «о» представлено на ней только коротким сег-

ментом, создающим скорее ощущение призвука, чем полноценного аллофона. Тем не менее, этот сегмент присутствует, хотя его выпадение не привело бы ни к появлению иного замечательного слова, ни к ошибочной идентификации данного слова при обращении к перцептивному словарю.

В данный момент я вижу следующие варианты возможного объяснения этого явления, каждый из которых требует проверки на большем материале.

Первый вариант — чисто фонетический. Предударный гласный после сегмента «не» может выпадать, если за ним следует кластер согласных. В нашем случае — «бх», «тд» ([пх] и [дд] в фонетической транскрипции).

Второй вариант. Слова *необходимый* и *неотделимый* ни при каких условиях не могут интерпретироваться как произнесенные с отрицательной частицей слова *обходимый* и *отделимый*, поскольку последние не существуют в современном русском языке. Иная ситуация со словами *неокрашенный* и *необоснованно*. Первое в звучащей речи может быть реализацией или прилагательного, или причастия с отрицательной частицей. Второе — наречия или краткого причастия. (Здесь надо отметить, что двойной сонант в наречии *необоснованно* на спектрограмме рис. 2 оказался короче одиночного в предшествующем слого.) Решение же о том, произнесено одно слово или два (отрицательная частица + причастие), может быть принято только на этапе грамматического анализа целого высказывания. Можно предположить, что рассматриваемый сегмент сохраняется в речи (и в перцептивном словаре) ради снятия лишней неопределенности на следующих стадиях анализа высказывания.

Литература

1. Венцов А.В., Риехакайнен Е.И., Слепокурова Н.А. Ментальный лексикон и восприятие редуцированных форм // Труды междисциплинарного семинара «Анализ разговорной русской речи» (АРЗ-2007).- СПб.: ГУАП, 2007.- С. 21-24.
2. Венцов А.В. Восприятие устной речи и ментальный лексикон // Русская языковая личность: Материалы шестой выездной школы-семинара.- Череповец: ГОУ ВПО "Череповецкий государственный университет", 2007.- С. 63-69.
3. Горбова Е.В. и др, Предварительные результаты мониторинга современной русской спонтанной речи // Современная русская речь: Состояние и функционирование. Вып. II.- Под. Ред. С.И. Богданова, Н.О. Рогожиной и Е.Е. Юркова. – СПб. Филол. Ф-тет СПбГУ, 2006.- С. 7-31.

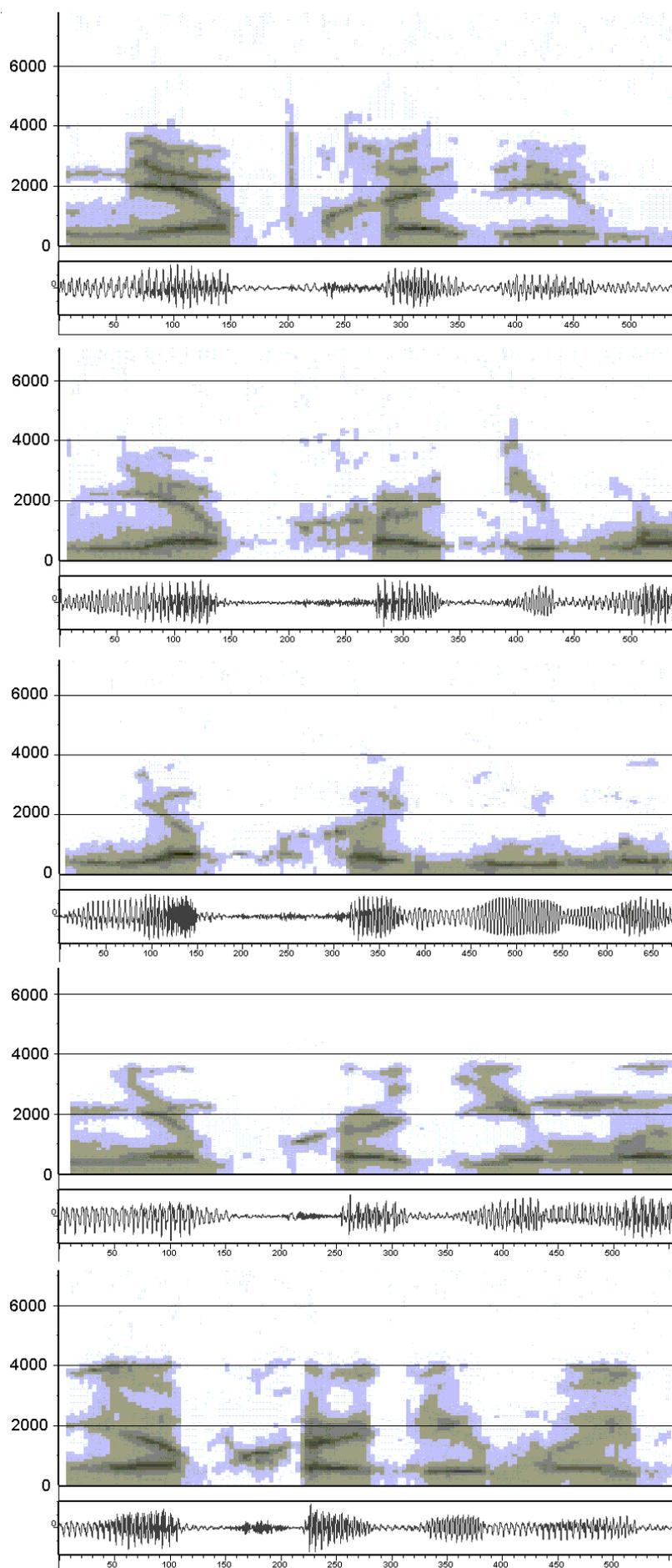


Рис. 1. Динамические спектрограммы слов *необходим...*

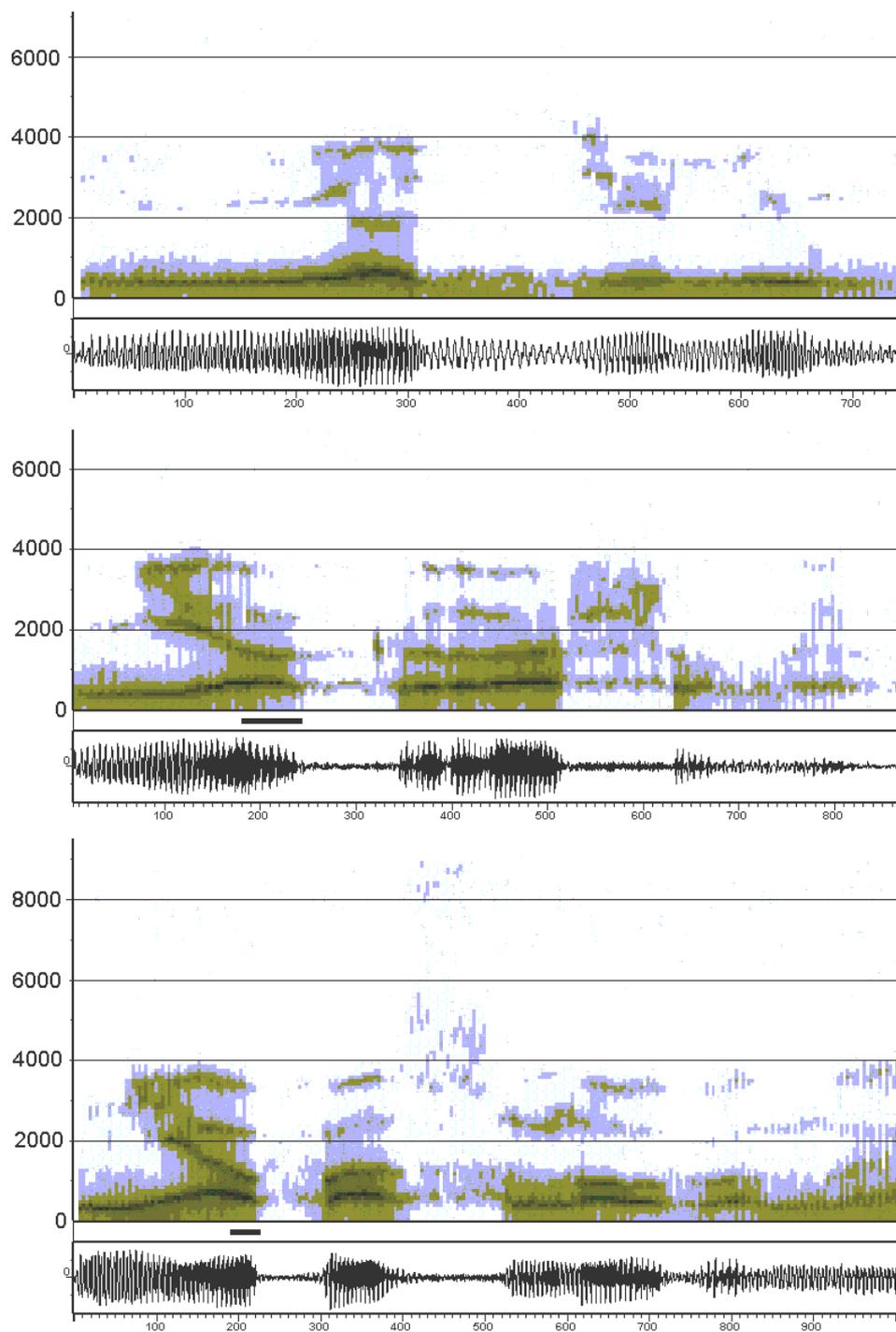


Рис. 2. Динамические спектрограммы слов (сверху вниз) *неотделимы*, *неокрашенной*, *необоснованно*. Жирным штрихом под спектрограммами отмечены позиция и длительность сегмента, соответствующего предударному «о».

Обзор подходов к моделированию спонтанной речи

Санкт-Петербургский институт информатики и автоматизации РАН,

г. Санкт-Петербург, Россия

kipyatkova@iias.spb.su

1. Введение

Одной из проблем автоматического распознавания спонтанной речи является вариативность произношения. Одни и те же слова могут произноситься различными дикторами по-разному, кроме того, произношение одного и того же человека может меняться в зависимости от стиля и темпа речи. Это приводит к снижению точности распознавания речи. В России данной проблемой стали заниматься относительно недавно, хотя работы, описывающие фонетические явления русской разговорной речи, появились более 40 лет назад [1, 2, 3, 4]. Качественная и количественная редукция гласных, ослабление согласных, выпадение согласных, уменьшение степени контрастности между гласными и согласными в пределах слога являются основными особенностями спонтанной речи [3]. В монографии О.А. Лаптевой [5] подробно рассматриваются фонетические особенности телевизионной речи: редуцирование слов, особенности согласных, особенности гласных, явления словесного ударения, отдельные явления фразового ударения. В работе [6] представлены наиболее частые замены звуков в русской речи. Было показано, что только 75% гласных фонем реализуются в «своих» аллофонах. В статье [7] рассматриваются три основные проблемы, с которыми сталкиваются при создании систем автоматического синтеза русской речи: 1) неоднозначность перехода буква→звук; 2) вариативность, обусловленная влиянием соседних звуков; 3) позиционные и другие типы вариативностей. Также здесь предлагаются способы их преодоления.

В современных системах распознавания спонтанной речи для учета вариативности произношения, помимо базовых транскрипций, которые создаются по фонетическим правилам, в словарь включают список альтернативных транскрипций. От метода, выбранного для создания альтернативных транскрипций, будет зависеть, насколько точно они отражают вариативность разговорной речи. Существует два основных подхода к нахождению (описанию) вариативности произношения [8]: методы, основанные на знаниях, и методы, основанные на данных. Рассмотрим эти методы более подробно.

В методах, основанных на знаниях, вариативность произношения определяется путем анализа существующих фонетических и лингвистических знаний, сформулированных экспериментальной фонетикой в ходе анализа речевых данных и акустико-артикуляторных характеристик фонем. Главная проблема этих методов заключается в том, что они описывают только часто возникающие отклонения в речи, при этом учитывается достаточно узкий фонетический контекст. Генерация транскрипций производится путем применения всех возможных комбинаций правил к базовым транскрипциям, при этом обычно не учитывается частота встречаемости правил, а также их комбинаций. Поэтому в зависимости от количества используемых правил синтезируется слишком много или слишком мало вариантов произношения.

В методах, основанных на данных, альтернативные транскрипции находятся в процессе анализа корпусов спонтанной речи. Собранные реальные транскрипции слов могут описать только те отклонения, которые встретились в данном корпусе. Поэтому полнота альтернативных транскрипций будет напрямую зависеть от размера и представительности речевого корпуса. В отличие от методов, основанных на знаниях, здесь по обучающему корпусу можно посчитать вероятности появления каждой альтернативной транскрипции и частоту возникновения каждого типа отклонения. Тем не менее, если корпус недостаточно представлен, то полученные альтернативные транскрипции, а также их вероятности могут быть ха-

рактельны для некоторого частного случая и не смогут описать все возможные отклонения, возникающие в спонтанной речи.

Для обоих подходов различают прямое и косвенное моделирование. В методах, основанных на знаниях, прямое моделирование осуществляется путем последовательного анализа каждой базовой транскрипции и добавления альтернативных вариантов произношения с учетом знаний эксперта. При косвенном моделировании используется некоторый набор правил редукции и ассимиляции, которые применяются для синтеза альтернативных транскрипций по имеющемуся списку базовых транскрипций.

При прямом моделировании в методах, основанных на данных, в качестве альтернативных транскрипций используются только часто встречающиеся в обучающем корпусе варианты произношения слов. При косвенном моделировании выявляются наиболее характерные изменения в произношении одинаковых цепочек фонем в различных словах, то есть по речевому корпусу определяют правила наиболее типичных изменений на уровне фонем.

Для обоих подходов при выборе метода косвенного моделирования необходимо осторожно обобщать наблюдаемую вариативность. Если некоторая вариативность появляется в сильно различающихся контекстах, обобщение может быть неправомерным. Так, наблюдаемая вариативность в служебных словах не обязательно подходит для ключевых слов, даже если фонетический контекст является тем же. Например, служебное слово “for” с канонической транскрипцией [f a o r]¹ имеет альтернативную транскрипцию [f e r]. Однако подобное преобразование неправомерно для существительного “forest” с канонической транскрипцией [f a o r a h s t].

Чтобы избежать появления некорректных правил в подходе, основанном на данных, анализируются два корпуса речи: обучающий и тестовый. В ходе косвенного моделирования вариативности транскрипций выбираются только те правила изменения произношения, которые были выявлены в обучающем корпусе и сохранились / подтвердились при анализе тестового корпуса речи.

Существует ряд причин, по которым применение косвенного моделирования в подходе, основанном на данных, является более перспективным. Во-первых, списки слов, использованных в обучающем и тестовом корпусах, могут отличаться. Поэтому альтернативные транскрипции для слов, которые присутствуют в обучающем корпусе, но отсутствуют в тестовом корпусе, будут отклонены. Использование правил позволяет описать наиболее характерные изменения, которые наблюдаются в ряде слов, что обеспечивает сохранение правил в процессе их проверки на тестовом корпусе. Во-вторых, правила строятся для цепочек фонем, а не полных транскрипций, и поэтому одно и то же правило может выполняться в различных словах. Так как статистика по встретившимся правилам будет более представительной, чем статистика по альтернативным транскрипциям, то и достоверность правил будет более высокой. Наконец, прямое моделирование межсловных явлений не представляется возможным, поскольку это требует полного перебора всех возможных комбинаций транскрипций слов. В случае косвенного моделирования возможные изменения на стыке слов могут быть заданы с помощью правил, которые применяются для любых комбинаций слов. В некоторых случаях для часто употребляемых пар знаменательных и служебных слов создаются базовые транскрипции, в которых уже учтены межсловные изменения на стыке служебного и знаменательного слова [10]. Например, для цепочки слов “going to” обычно сразу создается базовая транскрипция [g a a n a x].

Среди первых подходов к моделированию правил произношения можно выделить метод, использующий решающие деревья (Classification and regression trees) [11, 12]. Для каждого звука создавалось собственное дерево решений. При анализе обучающего корпуса выявленные замены звуков учитывались в листьях деревьев. При этом вероятность решения (то есть перехода в конкретный лист) зависела от числа встретившихся таких замен в обучающем корпусе.

¹ Здесь и далее используется «фонетический алфавит» Carnegie Mellon University (CMU) [9]

В последнее время наблюдается тенденция к переходу от методов, основанных на знаниях, к методам, основанным на данных. Так, в [13, 14] было показано, что основной целью лексических моделей не является описание спонтанной речи. Только 33% транскрипций, найденных в той части Switchboard, которая была размечена вручную, присутствовало в использованном словаре Pronlex. Ошибка распознавания слов с неканоническими транскрипциями оказалась на 11% больше, чем при распознавании слов, произнесенных в соответствии с каноническими транскрипциями. Другим наблюдением было то, что большую вариативность имеют те сегменты (цепочки фонем), которые чаще встречаются в речи. Причина распространности методов, основанных на данных, заключается в том, что они ориентированы на моделирование часто встречающихся сегментов. В тоже время, более детальное описание таких сегментов позволит существенно повысить точность распознавания речи, так как слова, употребляемые чаще всего, будут иметь большее влияние на уровень ошибок слов. Также следует заметить, что большинство современных систем автоматического распознавания речи основано на статистических подходах, и различия и сходства звуков, воспринимаемые людьми, не всегда являются полезными для автоматических систем. Поэтому обычно используют комбинацию методов, основанных на знаниях и данных. Например, с использованием знания о явлениях редукции и ассимиляции, составляется некоторый набор правил, по которому синтезируется словарь, включающий альтернативные транскрипции. Затем по размеченному вручную речевому корпусу проверяется, какие из альтернативных транскрипций действительно существуют, и оцениваются вероятности соответствующих правил редукции и ассимиляции.

При создании словаря альтернативных транскрипций обычно используется несколько итераций. При нахождении новой транскрипции требуется переобучение акустических моделей путем повторного транскрибирования обучающего набора с использованием новой транскрипции. Обычно это приводит к увеличению точности распознавания. Последующая итерация с использованием переобученных моделей для новых транскрипций позволяет выявить новый набор транскрипций, что может дать еще большее увеличение в производительности. Моделирование вариативности произношения на уровне лексических моделей уменьшает необходимость моделирования вариативности в акустических моделях. Переобучение акустических моделей после модификации словаря транскрипций дает лучший результат распознавания из-за того, что модели становятся более ориентированными.

2. Методы, основанные на знаниях

Использование лингвистических правил преобразования фонем для создания альтернативных транскрипций с помощью решающих деревьев было описано в [15]. Для того чтобы выбрать среди правил те, которые присутствовали в обучающем корпусе Switchboard (с использованием диктороадаптированного распознавателя), был использован алгоритм принудительного выравнивания, который обеспечивает оптимальное совмещение акустических моделей фонем и слов по речевому сигналу [16]. Затем выбранные правила были описаны в виде решающих деревьев с учетом контекстной зависимости от фонетических соседей, типов слов, скорости речи, средней продолжительности слов / звуков, ударности гласных и вычисленных вероятностей переходов по дереву. Вероятность правила оценивалась, исходя из относительной частоты использования каждого правила. Результирующие правила преобразования интерпретировались как зависимые от стиля произношения. Словарь был расширен за счет транскрипций, найденных с помощью принудительного выравнивания, то есть было проведено прямое моделирование, но на промежуточном шаге использовалось косвенное моделирование. Однако при использовании расширенного словаря точность распознавания не увеличилась. Авторы объяснили это увеличившейся неоднозначностью вследствие появления новых близких по звучанию транскрипций. Только после использования весов для каждой альтернативной транскрипции, зависящих от стиля произношения, уровень ошибок уменьшился на 7%. В этом подходе за основу был взят метод, основанный на знаниях, но ре-

чевой корпус использовался для верификации правил преобразования канонических транскрипций.

В статье [17] используется косвенный метод, основанный на знаниях, учитывающий вариативность произношения для тайского языка за счет применения четырех групп правил контекстной замены фонем. Вариативность произношения учитывается с помощью создания деревьев решений. Первая фонема каждой транскрипции помещается в корневой узел решающего дерева. В зависимости от следующей фонемы в дереве строится несколько вариантов изменения текущей фонемы. Если рассматривается не корневой узел, то на изменение фонемы влияет как предыдущая, так и последующая фонемы. Вероятности дуг решающих деревьев настраиваются с использованием алгоритма Витерби при распознавании обучающего корпуса речи. Также в этой работе была использована методика обучения акустических моделей [18] с автоматическим изменением разметки речевых сигналов. Суть методики сводилась к тому, что на начальном этапе обучения использовались канонические транскрипции, а по мере выявления случаев об изменении фонем в транскрипциях и формировании деревьев решений строились новые транскрипции, и они же использовались при новой разметке обучающего речевого корпуса. Размеченный с учетом новых транскрипций корпус снова использовался для обучения акустических моделей. Процесс обучения продолжался итерационно, пока точность распознавания речи не достигала необходимого уровня. Применение данной методики и учет вариативности произношения за счет деревьев решений позволили увеличить точность распознавания речи на 3,31%.

Также основанный на знаниях метод был применен в [19], где описывается метод распознавания слитной речи, использованный в диалоговой системе OVIS. Альтернативные транскрипции, содержащие внутрисловные замены, генерировались путем применения фонологических правил. Пять правил произношения для голландского языка были исследованы в [20] (четыре правила редукции и одно правило вставки). Было показано, что имеет смысл объединять эти правила в определенных контекстах. При этом межсловная вариативность произношения моделировалась двумя способами: 1) путем добавления межсловных вариантов произношения в словарь; 2) путем добавления в словарь цепочек слов и их альтернативных вариантов. Второй способ показал несколько лучший результат распознавания. После добавления альтернативных транскрипций в словарь производилось переобучение моделей фонем с помощью автоматического транскрибирования обучающего корпуса с использованием распознавания слитной речи в режиме принудительного выравнивания. Лучший результат был получен при объединении альтернативных транскрипций, содержащих внутрисловные замены, и транскрипций с учетом межсловной вариативности, сформированных с помощью второго способа. По сравнению с базовой системой точность распознавания увеличилась на 8,8%.

3. Методы, основанные на данных

Для нахождения оптимального набора альтернативных транскрипций в [21] был использован прямой метод, основанный на данных. Для экспериментов использовалась речевая база данных американского английского со словарем, содержащим 991 слово. В корпусе были отмечены только границы слов, а в словарь вошли их фонетические транскрипции. Никакие лингвистические и фонетические правила, позволяющие учесть вариативность произношения, не использовались на этапе обучения. При помощи фонетического декодера был произведен подбор лучших гипотез распознавания (цепочек фонем) для каждого слова в речевом корпусе. Затем в ходе оптимизации числа альтернативных транскрипций и настройки акустических моделей фонем на основе критерия максимума правдоподобия был окончательно сформирован словарь системы распознавания. Авторам удалось повысить точность распознавания слов от 6 до 19% в зависимости от сложности акустических моделей фонем.

Другой метод нахождения альтернативных вариантов произношений слов по речевому корпусу был представлен в [22, 12]. Вначале с помощью фонетического декодера, учитывающего биграммы фонем, были получены альтернативные транскрипции. Затем отклоне-

ния, обнаруженные в альтернативных транскрипциях, были представлены в виде деревьев решений (правил). После этого обучающий набор фраз был повторно транскрибирован с использованием этих правил для получения «сглаженных транскрипций». В словарь добавлялись только те варианты произношения, которые встречались достаточно часто. Дополнительной особенностью этого подхода является возможность формирования «динамического» словаря в процессе распознавания речи. В этом случае словарь содержит не только транскрипции, но и информацию о частоте употребления слова, вероятность слова в модели языка, длине слова, темпе произнесения и другие параметры, которые полезны при моделировании произношения. В результате для текущей гипотезы слова подбираются не все его альтернативные транскрипции, а лишь некоторое подмножество, которое удовлетворяет данным параметрам. При тестировании системы распознавания спонтанной речи на корпусе SPRACH 98 уровень ошибок сократился на 1,4%.

Подобный подход был использован в проекте Verbmobil [23]. Посредством распознавания фонем создавались альтернативные транскрипции. В словарь были включены только те транскрипции, которые встретились более 7 раз в обучающем корпусе. Было обнаружено, что для слов, встретившихся 15 раз и более, был достигнут некоторый стабильный набор альтернативных транскрипций, который уже не изменялся с увеличением речевого корпуса. Другими словами, экспериментально было показано, что слово не может иметь более 15 альтернативных транскрипций, однако это значение может варьироваться в зависимости от прикладной области.

Примером косвенного моделирования методом, основанным на данных, являются эксперименты, проведенные в [24] на речевом корпусе британского английского для моделирования акцента. В этой работе не использовалась ручная разметка транскрипций, поэтому данный подход можно считать полностью основанным на данных. Преобразование гласных является наиболее важным отличием в акцентах британского английского. Поэтому при формировании альтернативных транскрипций были модифицированы акустико-лексические модели базовых транскрипций таким образом, чтобы допустить возможность замещения любой гласной. Затем в процессе принудительного выравнивания было отобрано по три лучших транскрипции для каждого слова, которые использовались для извлечения правил в виде контекстно-зависимых деревьев решений. При добавлении нового акцента заново формировался словарь транскрипций. В среднем по 4 транскрипции на слово было достаточно для описания вариативности речи, возникающей вследствие акцента. Тестирование было выполнено на словаре, содержащем 2000 слов, и добавление альтернативных транскрипций позволило увеличить производительность системы распознавания речи.

В работе [25] модель, построенная с учетом психолингвистических знаний о восприятии речи человеком, была модифицирована для распознавания спонтанной речи на голландском языке. Здесь также был применен прямой метод создания альтернативных транскрипций на основе фонетического декодера и принудительного выравнивания канонических транскрипций по полученным цепочкам фонем (реальным реализациям произношения слов). Выявленные таким образом правила изменения фонем были описаны с помощью решающих деревьев. В словарь добавлялись альтернативные транскрипции с априорной вероятностью выше 0,063. В результате получилось в среднем 77 вариантов на слово. Затем автоматический распознаватель речи, содержащий словарь всех альтернативных транскрипций, но обученный только на канонических транскрипциях, был запущен в режиме принудительного выравнивания, для того чтобы найти наиболее вероятный вариант произношения для каждого слова в обучающем корпусе. На последнем шаге словарь распознавания речи включал только наиболее вероятные варианты транскрипций, найденные при принудительном выравнивании. Для тестового материала проводилась подобная процедура по созданию словаря транскрипций, которые более точно описывают акустический сигнал. Затем происходило переобучение акустических моделей. Тестирование проводилось, используя 10643 фразы из корпуса DDAC, которые не использовались при обучении. Эти фразы содержали различные элементы, характерные для спонтанной речи, такие как озвученные паузы, незавершенные

слова, повторяющиеся слова и др. В отличие от базовой системы распознавания речи, которая показала 64,45% ошибок слов, расширение словаря позволило понизить уровень ошибки до 48,2%.

4. Комбинированные методы

Сравнение комбинированного метода Bottom-Up с методом Top-Down, основанном на знаниях о четырех основных редукциях в голландском языке, было проведено в [26]. В методе Top-Down словарь содержал базовые транскрипции уже с учетом всех возможных редукций. При формировании альтернативных транскрипций в методе Bottom-Up брались как варианты транскрипций с учетом редукции, так и полные варианты произношения. Затем за счет применения некоторого обучающего корпуса отбирались только те альтернативные транскрипции, которые лучше всего соответствовали речевому сигналу. В результате анализа оставшихся альтернативных транскрипций был сформирован набор соответствующих правил редукций, а также относительная и абсолютная частота их появления. Следует отметить, что при формировании правил учитывалась не только фонема, подверженная редукции или замене, но и ее фонетический контекст, поэтому число правил увеличилось по сравнению с методом Top-Down. На этапе распознавания речи были использованы только те правила, которые встретились в обучающем корпусе более 100 раз. Оба метода дали примерно одинаковую точность распознавания, однако в методе Bottom-Up использовался словарь меньшего размера. При этом альтернативные транскрипции, полученные при том и другом методе, совпадали на 96%.

В [27] представлена система ANGIE, генерирующая обобщенную вероятностную модель для альтернативных фонетических реализаций в разговорной речи. Применяя контекстно-зависимые фонологические правила к базовой транскрипции, определяются альтернативные варианты произношения. При выставлении вероятности альтернативным транскрипциям принимались во внимание такие факторы, как слоговая позиция, ударение, фонетический контекст, морфологические характеристики словоформ. Каждое слово в обучающем корпусе разбиралось, используя контекстно-независимую грамматику. Вероятностная модель получалась с помощью выявления характерных реализаций, обнаруженных в большом обучающем корпусе. При этом вероятностная модель накладывается на дерево разбора и обучается так, чтобы учитывать наиболее частые отклонения в произношении. Поскольку эти отклонения учитываются в виде правил, то они смогут покрыть вариативность слов, встретившихся в обучающем корпусе, а также всех других слов, для которых характерны подобные типы отклонений. Этот момент важен, потому что те правила, которые были сформулированы по ограниченному обучающему корпусу, могут быть обобщены для других слов, не встретившихся при обучении, но имеющих подобный слоговой контекст. В большинстве случаев это позволяет произвести обучение на речевых данных из одной области, а тестировать систему распознавания на данных из другой области, где доступного обучающего материала может быть недостаточно. При обучении модели использовались речевой корпус Jupiter, содержащий фразы по теме «прогноз погоды», а также небольшой набор речевого материала из корпуса Mercury по авиационной тематике. При этом при тестировании системы распознавания речи использовались только фразы из корпуса Mercury. Применение системы ANGIE для получения альтернативных фонетических реализаций позволило уменьшить ошибку распознавания слов на 5,8%, а на этапе понимания уровень ошибок сократился на 12,6% как для речи, записанной в тихих условиях, так и при наличии шума.

В работе [28] использовался метод, основанный на знаниях, в комбинации с принудительным выравниванием по алгоритму Витерби. Для слов, которые часто встречались в обучающем корпусе, веса транскрипций оценивались путем определения относительной частоты появления вариантов произношения. Для слов, которые нечасто встречались в обучающем корпусе, были применены правила, описанные в [29, 30]. Веса транскрипций переоценивались, используя решающие деревья. Вероятность применения каждого правила оценивалась по следующему набору критериев: (1) сколько раз созданная по правилам транскрипция

встречалась в обучающих данных; (2) контекстная зависимость фонем; (3) сколько раз слово наблюдалось в обучающих данных; (4) длина транскрипции слова по количеству фонем. Расширенный таким образом словарь был протестирован при распознавании фраз из корпусов Switchboard и Callhome и показал небольшое улучшение по точности распознавания слов.

5. Заключение

Анализ существующих подходов к моделированию вариативности спонтанной речи позволяет сделать следующие выводы. При использовании методов, основанных на знаниях, альтернативные транскрипции генерируются путем использования правил замен и редукций фонем внутри слова и на стыке двух слов. Применение комбинации правил в одном слове часто приводит к необоснованно большому числу транскрипций и появлению неправдоподобных вариантов. В результате сильно возрастает лексическая неоднозначность.

Методы, основанные на данных, позволяют создать дополнительные транскрипции путем анализа корпуса обучающих речевых данных. При недостаточном количестве обучающих данных существует риск, что сформированный набор альтернативных транскрипций не покрывает всех вариантов произношения, возникающих в реальной разговорной речи. В отличие от прямых методов, которые описывают вариативность каждого слова в отдельности, косвенные методы, описывающие преобразование фонем, могут быть применены для любого словаря, однако следует учитывать достаточно широкий фонетический контекст, чтобы избежать появления несуществующих транскрипций.

Для учета степени вариативности произношения слов в зарубежных исследованиях была введена мера неоднозначности (“confusability”, “consistency”, “consolidation” в разных источниках), характеризующая степень расширения словаря вследствие добавления альтернативных транскрипций.

В методах, основанных на данных, в большинстве случаев используется фонетический декодер, обеспечивающий автоматическую разметку фонем в речевом корпусе. Поэтому от точности работы декодера зависит, насколько точно будут определены реальные транскрипции слов. В то же время, для обучения самого фонетического декодера также необходим достаточно представительный корпус спонтанной речи с предварительно размеченными границами фонем. По этой причине обычно используется несколько итераций для переобучения моделей фонем и поиска новых альтернативных транскрипций по тому же обучающему речевому корпусу.

Проблема вариативности произношения слов русского языка давно исследуется лингвистами, а также специалистами в области автоматического синтеза речи. Предложенные методы в основном описывают некоторые закономерности редукции и ассимиляции звуков, возникающих в разговорной речи, которые можно формализовать в виде правил для создания альтернативных транскрипций. Автором был разработан модуль генерации альтернативных транскрипций по расширенным правилам транскрибирования [31]. Для проверки полученных транскрипций на обучающем речевом корпусе планируется применить комбинированный подход. Работа проводится при поддержке гранта РФФИ, проект №08-08-00128 “Моделирование нефонемных речевых элементов и создание альтернативных транскрипций для распознавания спонтанной русской речи”.

Литература

1. Русская разговорная речь. Под редакцией Е.А.Земской. М.: Наука, 1973. 485 с.
2. Гейльман Н.И. Разговорная речь // Проблемы и методы экспериментально-фонетического анализа речи/под ред. Л.Р. Зиндера и Л.В. Бондарко. Л.: Изд-во СПбГУ, 1980. С.110-122.
3. Светозарова Н.Д. Некоторые особенности фонетики русской спонтанной речи//Бюллетень фонетического фонда русского языка №8, Фонетические свойства русской спонтанной речи. СПб: Бохум, 2000. С. 7-15.
4. Кузнецов В.И. Вокализм связной речи. СПб.: изд-во СПбГУ, 1997. 247с.
5. Лаптева О.А. Живая русская речь с телеэкрана: разговорный пласт телевизионной речи в нормативном аспекте. М., 2007.

6. Скредин П.А. Формальные методы анализа речи: проблемы интерпретации результатов. Материалы XXXVI международной филологической конференции. СПб. 2007. С. 3-14.
7. Захаров Л.М. Акустическая вариативность звуковых единиц в русской речи // Язык и речь: проблемы и решения. Сборник научных трудов к юбилею профессора Л. В. Златоустовой / Под ред. Г. Е. Кедровой и В. В. Потапова. М.. 2004.
8. Amdal I. Learning pronunciation variation. A data-driven approach to rule-based lexicon adaptation for automatic speech recognition. PhD thesis. Department of Telecommunications Norwegian University of Science and Technology. Norway. 2002.
9. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
10. Fosler-Lusser E., Williams G. Not just what, but also when: Guided automatic modeling of Broadcast News. Proc. DARPA Broadcast News Workshop, Herndon (VA), USA, 1999. pp. 171-174.
11. Bahl L.R., Souza de P.V., Gopalakrishnan P. S., Nahamoo D., Picheny M.A. Decision trees for phonological rules in continuous speech. Proc. ICASSP-91, Toronto, Canada, 1991. pp. 185-188.
12. Riley M.D., Ljolje A. Automatic speech and speaker recognition: Advanced topics, ch. Automatic generation of detailed pronunciation lexicons. Kluwer. 1996. pp. 285-301.
13. Fosler-Lussier J.E., Morgan N. Effect of speaking rate and word frequency on pronunciations in conversational speech. Speech Communication. vol. 29. 1999. pp. 137-158.
14. Fosler-Lussier J.E. Dynamic pronunciation models for automatic speech recognition. PhD thesis. University of California. Berkeley. 1999.
15. Finke M., Waibel A. Speaking mode dependent pronunciation modeling in large vocabulary conversational speech recognition. Proc. EUROSPEECH-97, Rhodes, Greece, 1997. pp. 2379-2382.
16. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2). 1989. pp. 257–285.
17. Kanokphara S., Tesprasit V., Thongprasirt R. Pronunciation Variation Speech Recognition Without New Dictionary Construction. NECTEC Technical Journal, 13(4). 2003. pp. 317-321.
18. Tarsaku P., Kanokphara S., A Study of HMM-based automatic segmentations for Thai Continuous Speech Recognition System, Proc. SNLP2002.
19. Kessens J.M., Wester M., Strik H. Modeling Within-word and Cross-word Pronunciation Variation to Improve the Performance of a Dutch CSR. Proc. of the 14 th Int. Congress of Phonetic Sciences, San Francisco, 1999. pp. 1665 - 1668.
20. Kessens J.M., Wester M., Strik H. Improving the performance of Dutch CSR by modeling within-word and cross-word pronunciation variation. Speech Communication. vol. 29. 1999. pp. 193-207.
21. Holter T., Svendsen T. Maximum likelihood modeling of pronunciation variation. Speech Communication. vol. 29, 1999. pp. 177-191.
22. Fosler-Lusser E. Multi-level decision trees for static and dynamic pronunciation models. Proc. EUROSPEECH-99, Budapest, Hungary, 1999. pp. 463-466.
23. Wolf M., Eichner M., Hoffmann R. Automatic learning and optimization of pronunciation dictionaries. Proc. ISCA ITRW Adaptation method for speech recognition, Sophia-Antipolis, France, 2001. pp. 159-162.
24. J. J. Humphries, P.C. Woodland, D. Pearce. Using accent-specific pronunciation modeling for robust speech recognition. Proc. ICSLP-96, Philadelphia (PA), USA, 1996. pp. 2324-2327.
25. Scharenborg O., Boves L. Pronunciation Variation Modelling in a Model of Human Word Recognition. Proceedings of the Workshop on Pronunciation Modeling and Lexicon Adaptation, Estes Park, CO, USA, 2002. pp. 65-70.
26. Kessens J.M., Strik H., Cucchiaroni C. A bottom-up method for obtaining information about pronunciation variation. Proc. ICSLP-2000, Beijing, China, 2000. pp. I:274-277.
27. Seneff S. and Wang C. Modeling Phonological Rules through Linguistic Hierarchies. Proc. ISCA Workshop, Estes Park, CO. 2002. pp. 71-76.
28. Byrne B., Finke M., Khudanpur S., McDonough J., Nock H., Riley M., Saraclar M., Wooters C., Zavalagkos G. Pronunciation Modelling for Conversational Speech Recognition: A Status Report from WS97 IEEE Workshop on Speech Recognition and Understanding, Dec 14-17, Santa Barbara, California. 1997.
29. Ladefoged P. A Course in Phonetics. Harcourt Brace Jovanovich, Inc., New York, 1975.
30. Weintraub M., Fosler E., Galles C., Kao Y., Khudanpur S., Saraclar M., Wegmann S. Automatic Learning of Word Pronunciation from Data. 1996 LVCSR Summer Workshop Technical Reports, 1996.
31. Kipyatkova I. An algorithm of generation of alternative phonetic transcriptions for spontaneous Russian speech recognitions. Proc. of International forum “Information and communication technologies. Problems, perspectives”. 2008. pp. 45-48.

А.Б. Викторов, М.В. Ескевич

Адаптируемое под язык акустическое моделирование для распознавания речи

ООО «Одитек»
г. Санкт-Петербург, Россия
work@auditech.ru

1. Введение

Речь является наиболее удобным для человека способом передачи и получения информации. Передовые речевые технологии ориентированы на создание наибольших удобств для пользователя, что означает использование речевого общения на родном для пользователя языке. В настоящий момент локализации систем распознавания новых языков делаются с использованием десятков часов записанной и затранскрибированной речи. Однако создание подобной обучающей базы для каждого языка — это трудоемкий, дорогостоящий и длительный процесс. Количество языков, которые могут представлять потенциальный интерес, может достигать нескольких тысяч, но при этом их состав может меняться в зависимости от политической или экономической ситуации.

В данных условиях вызывает интерес разработка методики адаптации уже существующих акустических моделей для распознавания нового языка с использованием только ограниченного количества данных нового языка. В данной статье дается описание работы по созданию подобной системы адаптации.

2. Обобщенное множество единиц для транскрипции

Международный фонетический алфавит (МФА) [1] как система знаков для записи транскрипции на основе латинского алфавита, разработанной и поддерживаемой Международной фонетической ассоциацией, и системы его записи с помощью набора символов ASCII, такие как SAMPA [2], Worldbet [3], считаются основным способом представления фонетического материала. Однако исследователи, занимающиеся распознаванием речи, выбирают определенный набор этих символов или создают свой набор обозначения классов фонем или фонетических образов для решения конкретных задач.

Шульц и Вайбель пишут о разработке единого множества единиц для покрытия фонемного состава 12 языков (китайский, японский, корейский, турецкий, хорватский, испанский, английский, французский, немецкий, португальский, русский, шведский) на основе артикуляторной близости фонем и единообразия их обозначения через символы МФА (см. [4], [5], [6], [7]).

Для распознавания языка может использоваться инвентарь фонем, не зависящий от языка, полученный автоматически с помощью накапливающего алгоритма иерархической кластеризации. Коридор-Арно получил из 148 фонем 4 языков (английского, испанского, французского и немецкого) 83 кластера, из которых 48 представлены только в одном из языков, а 35 кластеров содержат фонемы из разных языков (см. [8]).

Как и в исследовании Шульца, для создания обобщенного множества единиц мы основываемся на артикуляторных характеристиках и единообразии и/или близости фонем/аллофонов в системе обозначения МФА, при этом добавляем отдельные символы для частотных аллофонов, которые используются при распознавании речи данных языков (например, разные степени редукции гласных в русском языке). Появляющиеся различия в артикуляторных характеристиках звуках разных языков, обозначенных одним символом МФА и объединенных в одну единицу множества, понимаются исследователями, но не рассматриваются как отрицательная статистика, потому что в дальнейшем для алгоритма распознавания эти различия могут моделировать вариативность произношения.

Языки, используемые нами для составления обобщенного множества единиц (русский, английский, немецкий, французский, испанский, итальянский, португальский, польский, словацкий, чешский, венгерский, арабский), выбраны исходя из имеющихся баз данных, пригодных для дальнейшей апробации системы распознавания, представляют собой языки разных языковых семей и являются распространенными.

Сформированное единое множество единиц состоит из 169 единиц, которые разделяются на полифонемы (фонемы, которые встречаются в двух и более языках) и монофонемы (фонемы, представленные только в одном из языков) (см. Таблицу 1.).

Таблица 1. Обобщенное множество единиц

Полифонемы, встречающиеся в 2-х и более языках																	
количество языков	количество фонем	согласные									гласные						
все	9	b	d	t	k	f	z	s	m	n							
11	5	p	ø	g	j						u:						
10	5	v	dʒ								à	o:	ý				
9	5	æ	r								a:	I	i:				
8	6	ts	í	ë	x						U	O					
7	3	dz	tʃ								e						
6	4	w	í	l	r:												
5	1	h															
4	17	b:	d:	d ^j	p:	t:	k:	m:	nj	n:	o	e:	á	au			
		s:	ø:	tʃ ^j	f:												
3	25	M	l:	É	v:	g:	Ã	ð	tø:	ts:	ÿ	œ	î	i	y	ia	ie
		t ^j	C	z:	ò	Ò					iu	è	aU	eu			
2	26	p ^h	t ^h	k ^h	dʒ ^j	dʒ:	Ð	z ^j	s ^j	v ^j	À	O~	Ý~	y:	io	wo	wɔ
		Æ	í:	ʃ							aI	ai	ou	oi	oI	we	wa
Итого полифонем:	106																
Монофонемы																	
язык	количество фонем	согласные									гласные						
EN	9	ð									Iâ	Uâ	OI	âr	O:	eI	eâ
											âU						
GE	3	R	в	pf													
FR	3	ц									î~	a~					
SP	1	ô															
IT	11	dʒ:	É:								iÝ	iɔ	wÝ	wi	ei	Ýi	Ýu
											Oi	ui					
PO	6										ß	ß~	í~	u~	e~	o~	
PL	1	dʒ															
CZ	4	r ₁									Ý:	eU	oU				
SLO	0																
HU	2	cç									î:						
AR	12	d ^o	t ^o	q	ö	s ^o	ã	ò:	ò:	Ê	aj						
		ö	l ^o														
RU	12	p ^j	b ^j	g ^j	f ^j	ù	x ^j	k ^j	g ^j	mj							
		r ^j									è	u					
Итого монофонем	63																
ВСЕГО единиц	169																

Для характеристики получаемого множества единиц используются следующие параметры (см. Таблицу 2.):

- количество единиц в языке
- количество единиц, специфических для данного языка
- коэффициент покрытия (процент единиц обобщенного множества, покрывающих алфавит данного языка)

Таблица 2. Характеристики языков.

	в среднем	RU	EN	DE	FR	SP	IT	PO	PL	SL	CZ	HU	AR
количество единиц в языке	50.4	74	44	50	37	46	74	40	39	50	39	63	49
количество единиц, специфических для данного языка	5.33	12	9	3	3	1	11	6	1	0	4	2	12
коэффициент покрытия	0.9	0.84	0.80	0.94	0.92	0.98	0.85	0.85	0.99	100	0.9	0.97	0.76
коэффициент полифонемности	3.58												

(RU — русский, EN — английский, DE — немецкий, FR — французский, SP — испанский, IT — итальянский, PO — португальский, PL — польский, SL — словацкий, CZ — чешский, HU — венгерский, AR - арабский)

Дополнительной характеристикой множества языков, использующихся для создания базы системы распознавания, является коэффициент полифонемности языков (K_{Π}): отношение между суммой количества фонем, присущих языкам, (N_i) и количеством единиц обобщенного множества (см. (1.)). Данная характеристика показывает среднее количество языков, которые включают в себя единицу обобщенного множества. У Шульца при создании единого множества единиц на основе 12 языков, частично отличающихся от выбранного нами набора, коэффициент полифонемности составил 2.99, то есть в среднем каждая единица единого множества встречалась в трех языках.

$$K_{\Pi} = \frac{N_1 + N_2 + \dots + N_{12}}{N} = \frac{605}{169} = 3.58 \quad (1)$$

3. Методика акустического моделирования и адаптации результатов для распознавания нового языка

Существуют различные подходы к моделированию акустических моделей при использовании данных нескольких языков. Фонема каждого языка обучается на данных только этого языка, то есть отсутствует совместное использование данных разных языков при обучении акустических моделей; при обучении акустических моделей может быть использован материал всех языков, а информация о языке-источнике может быть сохранена или нет.

Мы предполагаем реализацию модели смешанного акустического моделирования. При обучении акустических моделей обобщенного множества единиц будут использоваться данные всех языков. При этом во время обучения не будет сохраняться информация о языке, и само обучение будет проходить на материале всех языков. Получаемые в результате мультязыковые акустические модели не будут зависеть от языка, что позволит уменьшить сложность и количество параметров системы распознавания. Также при объединении материала разных языков для создания одной акустической модели в систему закладывается вариативность произношения, что, возможно, будет иметь большее влияние на результат работы системы распознавания, чем прописанные иным образом правила вариативности модели.

Следующим этапом в разработке является использование полученной системы распознавания для моделирования системы декодирования нового языка при небольшом количестве материала.

Проблемы, которые вызваны малым количеством данных нового языка:

1. сложность в определении акустических моделей, которые будут использоваться для моделирования нового языка
2. большое фонетическое различие между новым языком и имеющимися моделями, выявляющееся при расширении окна фонетического контекста для создания контекстно-зависимых фонетических моделей

Размеченный материал нового языка подается на вход имеющейся системы для распознавания, построенной на материале 12 языков. При этом выбирается подмножество обобщенного множества единиц, необходимое для описания фонетического строя данного языка.

Параллельно с адаптацией созданной системы для работы с новым языком, будет строиться система распознавания данного языка, чтобы в дальнейшем можно было произвести сравнение результатов работы двух систем.

4. Заключение

Нами была проведена работа по подготовке к созданию системы распознавания речи, основанной на материале нескольких языков. Для этого на основе артикуляторных характеристик и единообразия и/или близости фонем/аллофонов в системе обозначения МФА было выделено единое множество 169 единиц. В дальнейшем мы предполагаем провести кластеризацию материала и сравнить полученные классы с определенными раньше.

Следующим шагом станет создание системы распознавания на материале нескольких языков, рассегментированных на единицы обобщенного множества. Для адаптации системы распознавания нового языка из всего множества акустических моделей будет выбрано подмножество, необходимое для описания его фонемного состава. Работа системы, построенной на материале нескольких языков, будет сравниваться с результатами системы распознавания, обученной на материале только нового языка.

Литература

1. <http://www2.arts.gla.ac.uk/IPA/ipa.html>
2. <http://www.phon.ucl.ac.uk/home/sampa/index.html>
3. J. L. Hieronymus, ASCII Phonetic Symbols for the World's Languages: Worldbet
4. <http://www.anticline.ling.gu.se/~jimh/courses/ipa.ps>
5. T. Schultz and A. Waibel, Language Independent and Language Adaptive Large Vocabulary Speech Recognition, Proceedings of the International Conference of Spoken Language Processing (ICSLP-1998), Vol. 5 pp 1819--1822, Sydney, Australia, November 1998.
6. T. Schultz, Multilinguale Spracherkennung. Kombination akustischer Modelle zur Portierung auf neue Sprache. Zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften vorgelegte Dissertation, Universität Karlsruhe, June 2000.
7. T. Schultz and A. Waibel, Language Portability in Acoustic Modeling, Proceedings of the Workshop on Multilingual Speech Communication (MSC-2000), pp 59-64, Kyoto, Japan, October 2000.
8. T. Schultz and A. Waibel, Language-independent and language-adaptive acoustic modeling for speech recognition, Speech Communication, vol. 35, pp. 31--51, August 2001. Issues 1-2. <http://citeseer.ist.psu.edu/schultz01language.html>
9. C. Corredor-Ardoy, J.L. Gauvain, M. Adda-Decker, L. Lamel, Language identification with language-independent acoustic models, Proc. Eurospeech '97, Rhodes, Greece, pages 55--58, 1997, <http://citeseer.ist.psu.edu/18467.htm>

А.В. Вартапов

Психофизиологический подход к проблеме распознавания речи

*Московский государственный университет имени М.В. Ломоносова,
г. Москва, Россия
a_v_vartanov@mail.ru*

В системе человеческого Языка обыкновенно 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 двухплановая совокупность содержательных, собственно языковых средств, которая дальше распадается на тесно сплетенные, нечетко разграниченные подсистемы лексики и грамматики [7]. Звуковые оболочки различаются по трем уровням содержания как 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» [7]: морфема — это только "префабрикат" из фонологических единиц, озвучиваемый при вхождении в звуковую оболочку минимального автономного знака — слова. Звуковая оболочка слова не равна совокупности входящих в нее морфемных оболочек, так как она приобретает дополнительную характеристику — акцентную модуляцию (ударение). Точно так же не равна последовательности оболочек входящих в нее слов и звуковая оболочка фразы, которая приобретает дополнительную, не присущую словесным оболочкам фразовую интонационную модуляцию (что относится уже к сфере просодики). При этом существуют индивидуальные особенности голоса и его изменения в зависимости от состояния говорящего и от его отношения к произносимому, а также соответствующая эмоциональная экспрессия. Все это требует решения вопроса о комплексном подходе при разработке компьютерных систем распознавания естественной (слитной) речи, использования антропоморфных моделей анализаторов, интеграции в едином продукте одновременно действующих подсистем распознавания диктора, эмоционального тона, темпа, акцентной модуляции и других просодических переменных вместе с подсистемами распознавания языковых знаков всех уровней. В связи с этим представляется важным использовать модели из нейроподобных элементов, которые прошли верификацию в психофизиологических экспериментах. Так, для задачи учета проявления эмоций в звучащем слове можно рекомендовать четырехмерную сферическую модель эмоций [3, 4, 5, 6].

Фонемоцентризм классической фонологии и опирающиеся на нее наиболее распространенные модели компьютерного анализа и синтеза речи проявились в попытке практической реализации идеи о возможности и необходимости свести огромное, кажущееся неисчерпаемым многообразие звуков речи к небольшому, легко обозримому инвентарю в несколько десятков букв. Фактически, буква стала символом соответствующей фонемы как мельчайшей (неразложимой) базовой фонетической единицы. Эта идея — быстро свести (на основе различных алгоритмов сравнения) речевой звук (вариант произношения фонемы) к одному из звуковых эталонов (символу алфавита из большого набора), а уже затем анализировать лексический и грамматический уровни (используя большие словари и грамматические принципы образования словоформ) для корректировки первоначального выбора, на первый взгляд кажется самой простой и технологически удобной. Она положена в основу многих активно разрабатываемых программных продуктов — несмотря на то, что их эффективность очень далека даже от минимально достаточной для практических приложений. Однако дальнейшее развитие фонологической теории поставило вопрос о внутренней структуре фонемы. Наиболее актуальным из трех аспектов проблемы членимости фонемы (корме множественности реализаций — аллофонов и возможности расщепления фонемы слоговыми или морфемными границами) является вопрос о выделении в ее структуре собственно фонологических элементарных компонентов [7]. Уже в 1936 г. было показано, что фонемы неразложимы на более мел-

кие фонологические единицы, наделенные смысловоразличительной способностью, но не следующие друг за другом линейно, — например, звонкость, смычность. Однако нелинейная синтагматика — далеко не единственная особенность субфонемной элементарной единицы. Не менее важно и то, что компоненты фонемы, одновременно реализуемые в звуковой единице, сами по себе звуками не являются. Как отмечает В.Я. Плоткин [7], недооценка незвукового компонента звука связана с затянувшейся односторонностью как фонетики, так и фонологии. Но звуковой строй языка звуками вовсе не исчерпывается. Функция звука как продукта артикуляции и объекта аудиоперцепции — экстернализация речевого сигнала и перенос его от человека к человеку. Однако путь речевого сигнала начинается не во рту и завершается не в ухе — отправителем и получателем сигнала является мозг; это отмечается и в филологических трудах [7, 8]. Путь сигнала состоит из трех этапов: 1) от мозга к артикулирующим органам говорящего в виде нервного кода; 2) в виде звука к органу слуха воспринимающего; 3) в виде нервного кода к мозговым областям распознавания и понимания речи. И, как справедливо отмечает В.Я. Плоткин [7], к языковому плану выражения принадлежат обе субстантные реализации сигнала — как звуки, так и нервный код. Еще в начале XX века Бодуэн де Куртене определил язык как «слышимый результат правильного действия мускулов и нервов» [1, т.1, с.77]. Непризнание этого, выдвигание фонемы на центральную роль единственной базовой фонологической единицы неминуемо обрекло субфонемные фонационные компоненты на подчиненный статус признаков фонемы, средств для ее идентификации и различения. Так, Р.О. Якобсон определил фонему как пучок дифференциальных элементов [9], составив универсальный перечень «различительных признаков» из 12 оппозитивных пар, используемых всеми естественными языками. Однако, как отмечает В.Я. Плоткин [7], любые универсальные перечни мельчайших языковых единиц — будь то по Якобсону или по Н. Хомскому — неизбежно оказываются инвентаризациями звуко-различительных потенций человека, которые не преобразуются в готовом виде в первоэлементы языковых систем. Простое перечисление недостаточно для описания языковых единиц, так как на каждом уровне они закономерно объединяются в специфические для каждого языка парадигматические системы. Мельчайшие фонологические единицы характеризуются такими общими для любой языковой системы свойствами, как субстантная база (в смысле универсальной антропоники), парадигматические и синтагматические отношения между единицами, вариативность их реализации, функциональные отношения с вышестоящими единицами — как звуковыми, так и знаковыми. При этом субфонемные элементы, как и все языковые единицы, одновременно и дистинктивны, и содержат атрибут признаковости, поэтому в терминологическом плане обозначение «различительный признак» неудачно в обеих своих частях. Наиболее подходящее для них наименование — термин «кинакема», предложенный Бодуэном де Куртене и содержащий указание на две сопряженные сферы действия данной единицы — звукопроизводство (корень «кин») и звуковосприятие (корень «ак»), отграничивая ее от смежной единицы — фонемы, чья сфера действия (звук) лежит между ними. Существенная разница между статусом «различительного признака» и кинакемы, как подчеркивает В.Я. Плоткин [7], проявляется, помимо всего прочего, в том, что если перечни признаков видятся универсальными, то кинакемная система в принципе уникальна для каждого языка.

Как известно [7], звуковые оболочки всех знаковых единиц представляют собой участки звукового потока, который, естественно, подвергается сегментации и модулированию. В оболочках достаточно большой протяженности (фразовых, словесных) сегментация и модулирование определяются с учетом семантических факторов, на более же низком уровне звуковой поток организуется собственно фонологическими средствами без участия семантики. Элементарное модулирование потока осуществляется особой фонологической единицей — слогом, представляющим собой мельчайший, далее не членимый сегмент звукового потока. Внутренняя структура слога определяется принципом контраста между его компонентами, которые представляют собой хорошо различимые звуки. Стандартный квант звука — фонема. Звук может быть произведен лишь определенной совокупностью артикуляционных действий и воспринят как совокупность их перцептивных коррелятов. Компоненты таких сово-

купностей — кинакемы, кванты нервного кода аппарата звукопроизводства и звуковосприятия. Внутренняя структура фонемы может быть описана как блок кинакем. Фонемный блок кинакем — это стандартизованная, автоматизированная совместная реализация системно организованной совокупности кинакем. Принципиальная нелинейность, симультанность синтагматического объединения кинакем в фонемные блоки позволяет говорить о совместности реализации кинакем в блоке. Однако антропофонические и анатомо-физиологические факторы могут обусловить некоторое опережение или запаздывание в реализации тех или иных входящих в фонемный блок кинакем, в результате чего реализация запаздывающих кинакем предшествующей фонемы может накладываться на реализацию опережающих кинакем последующей фонемы, что затрудняет распознавание таких переплетающихся фонем [7].

Представляя собой команду для организации нервно-мышечной передачи, кинакемная парадигматика характеризуется тремя дихотомиями. Первая дихотомия обусловлена обязательной бинарностью кинакем: каждой кинакемной команде на то или иное артикуляторное действие закономерно противостоит команда на воздержание от этого действия, т.е. у каждой положительной команды есть антагонист в виде отрицательной команды (или отсутствия этой команды). Вторая фундаментальная дихотомия в кинакемных системах связана с внутрислоговой модуляцией речевого потока, разделяющей звуки на гласные и согласные с контрастными артикуляционными и перцептивными характеристиками: все кинакемные системы состоят из двух подсистем — вокалической и консонантной. Третья фундаментальная дихотомия делит все кинакемы на две категории — модальную и локальную: к модальной категории относятся кинакемы, управляющие способами образования звука и вертикальными параметрами речевого канала (осью сонорности по Jakobsonу), а к локальной — кинакемы, функционирующие в зоне регуляции горизонтальных движений (на оси тональности по Jakobsonу). Наряду с процессом максимального развертывания бинарного потенциала в кинакемных системах широко используется и другой путь обогащения их состава. Он заключается в привлечении добавочных оппозиций, субстантная база которых лежит за пределами минимальных кинакемных структур (это протяженность, палатализация, лабилизация, глоттализация, аспирация, назализация, абруптивность, скольжение растворное и тембровое, тоновая динамика и т.д.). Общее количество кинакемных оппозиций в одном национальном языке варьируется от 6-7 до 14-15.

Таким образом, кинакема как полноправная единица располагает своей, специфической для каждого языка парадигматической организацией, правилами внутрифонемной и межфонемной синтагматики. Она по-своему сегментирует и структурирует универсальную антропофоническую субстанцию артикуляционно-перцептивной деятельности, вступает в функциональные отношения с вышестоящей фонологической единицей — фонемой — и с морфемными оболочками, а в ряде случаев может даже самостоятельно выступать в роли подобной оболочки [7]. Специфичны для языков также способы разграничения слогов. Если фонему можно трактовать как блок кинакем, то вопрос о прохождении межслововой границы внутри фонемы можно ставить как вопрос о принадлежности входящих в фонему кинакем к разным слогам. В целом и сами кинакемные системы, и их элементы в каждом отдельном языке — кинакемы — уникальны как средства сегментации и организации универсальной антропофонической субстанции, формирования первичных квантовых единиц, функционирующих в составе конкретных звуковых и, шире, языковых систем. Уникальность каждой такой системы предполагает уникальность и формируемых ею элементов — языковых квантов, принципиально не сводимых в какие бы то ни было универсальные инвентари [7]. На рис. 1 представлена система кинакем и фонем русского языка из работы В.Я. Плоткина [7], где приведены такие системы для 100 различных языков.

Русский язык
Система кинакем

Подсистемы	Категории	Подкатегории	Оппозиции
Вокалическая	Модальная	Растворная	Узкорастворности Широкорастворности
	Локальная	Тембровая	Огубленности
Консонантная	Модальная	Преградная	Смычности Проточности
		Фональная	Шумности Сонорности
	Локальная	Активная	Предцентральнойности Зацентральнойности
		Пассивная	Зацентральнойности
Добавочная	Тембровая	Палатализации	

Система фонем

Гласные

-	0	+	Огубленность		
			Узкорастворность		Широкорастворность
i		u	+		-
e		o	+		+
	a		-		+

Согласные

	-			+					Палатализация				
+	-	-	-	0	+	-	-	-	0	Активная предцентральнойность			
-	-	-	+	0	-	-	-	+	0	Активная зацентральнойности			
0	-	+	0	0	0	-	+	0	0	Пассивная зацентральнойности			
										Смычн. Проточн. Шумн. Сонорн.			
p	t	k			p'	t'	k'			+	-	+	-
b	d	g			b'	d'	g'			+	-	+	+
m	n				m'	n'				+	-	-	+
f	s	ʃ	x		f'	s'	ʃ'	x'		-	+	+	-
v	z	ʒ			v'	z'	ʒ'			-	+	+	+
	l	j				l'				-	+	-	+
			ts					tʃ		+	+	+	-
			r					r'		+	+	-	+

Рис. 1. Система кинакем и фонем русского языка из работы В.Я.Плоткина [7].

Правомерность такого представления исследовалась в серии психофизиологических экспериментов по восприятию звуков речи русского языка. Эксперименты строились по методу многомерного шкалирования степени субъективного различия при восприятии пар стимулов. Стимулами служили звуковые образцы, записанные и предъявляемые испытуемым с помощью компьютера. В серии 1 исследовалось восприятие гласных звуков — использовалось 25 образцов (включая дифтонги естественные и искусственные, которые не используются в русском языке, но увеличивают статистическую надежность процедуры), из которых было образовано 300 сочетаний. Испытуемые оценивали степень различия предъявляемой пары образцов баллом от 0 (нет различия) до 9 (максимально сильное различие). Каждая пара предъявлялась по несколько раз в случайном порядке, оценки для каждой пары усредня-

лись (в том числе и по испытуемым, после специальной проверки на одинаковость их восприятия) и сводились в матрицу различий.

Далее эта матрица обрабатывалась метрическим методом многомерного шкалирования, что позволило оценить число различительных признаков данных звуковых образцов и построить пространственную модель восприятия гласных звуков.

Для этих экспериментальных данных было выделено 4 признака, которые после ортогонального вращения пространства получили следующую интерпретацию — широкорастворность, узкорастворность, огубленность и ось дифтонгов. На рис. 2 и 3 представлены проекции использованных звуковых образцов на оси 1-2 и 3-4 соответственно.

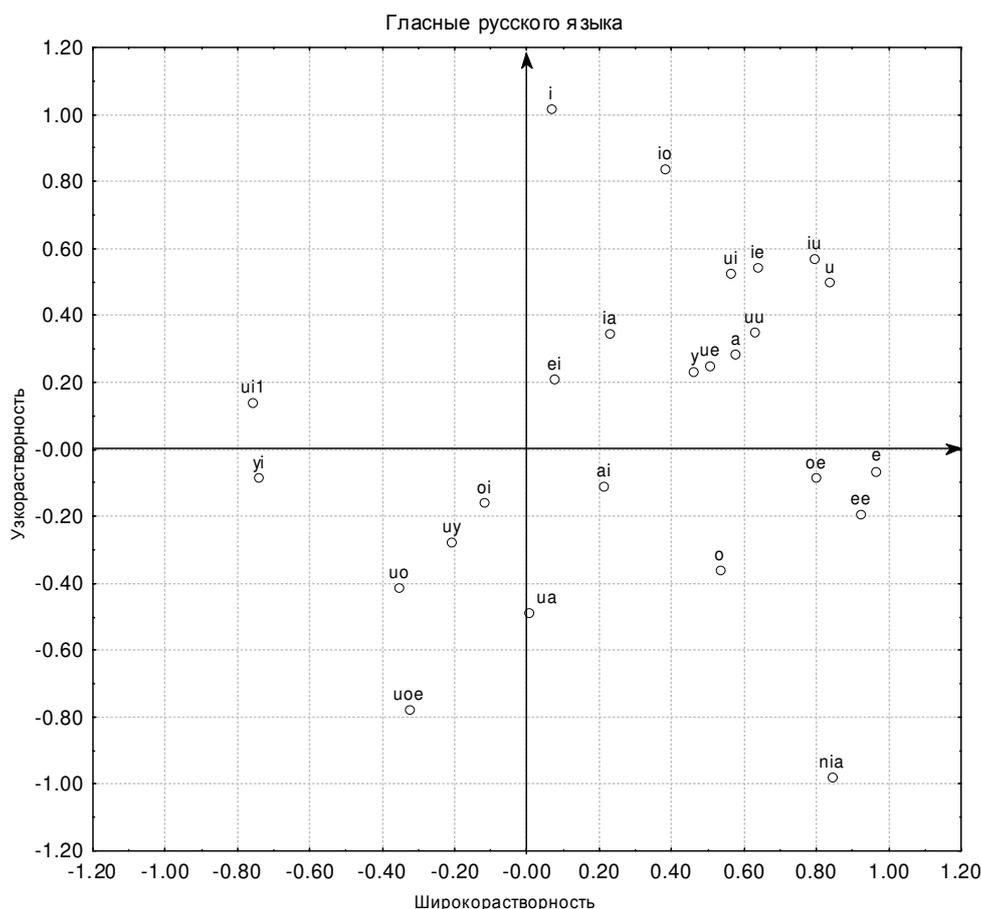


Рис. 2. Субъективное пространство восприятия гласных звуков русского языка в проекции на плоскость осей 1 и 2. Точки-стимулы подписаны условными транскрипциями (латинские буквы).

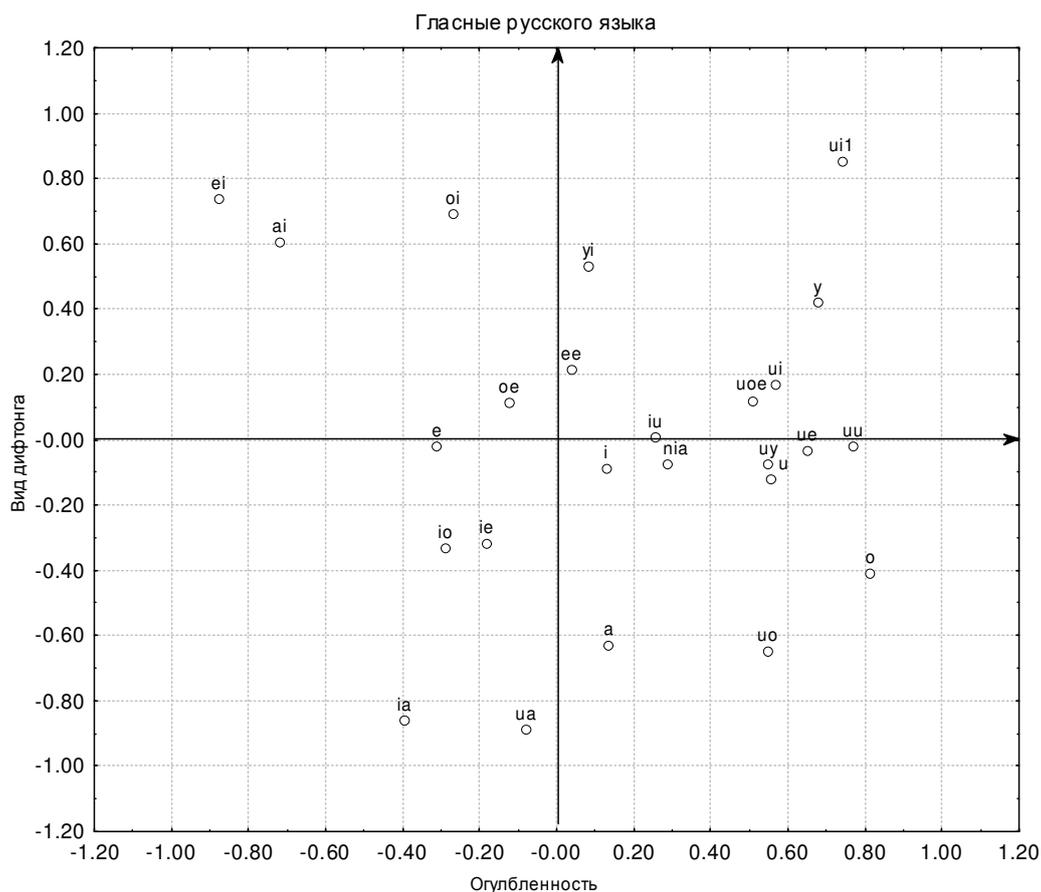


Рис. 3. Субъективное пространство восприятия гласных звуков русского языка в проекции на плоскость осей 3 и 4. Точки подписаны условными транскрипциями (латинские буквы).

Для проверки возможной интерпретации осей пространства субъективного восприятия человеком звучащих образцов гласных русского языка как кинакемых оппозиций была проведена вторая серия экспериментов. Задачами этой серии было:

- 1) Построение с помощью метода многомерного шкалирования субъективного пространства восприятия 14 гласных, при этом гласные «А» и «О» каждая присутствуют в 6 вариантах, различных по своим артикуляционным признакам произношения — широкорастворности и огубленности, соответственно (91 пара, предъявляемая не менее чем по 3 раза каждому испытуемому).
- 2) Выявление связи субъективных параметров восприятия гласных звуков и их объективных артикуляционных качеств — объективно измеренная (в см) степень широты раскрытия рта произносящего (для кинакемы широкорастворности) и степени огубленности.
- 3) Регистрация вызванных потенциалов мозга человека (ВП) на предъявление 6 выделенных стимулов (звуки «А», различающиеся произношением по артикуляционному признаку широкорастворность).
- 4) Сопоставление формы полученных ВП с помощью факторного анализа и выявление связи показателей ВП с субъективными параметрами восприятия гласных звуков.
- 5) Динамическая локализация мозговой активности на основе дипольной модели с помощью метода разделения ВП на корковую и глубинную составляющие [2].

Обнаружено, что для субъективного различения данного набора стимулов достаточно 3 признаков, которые могут быть интерпретированы как кинакемые оппозиции широкорас-

творности, узкостворности и огубленности. На рис. 4 представлено субъективное пространство восприятия в проекции на оси 1 и 2.

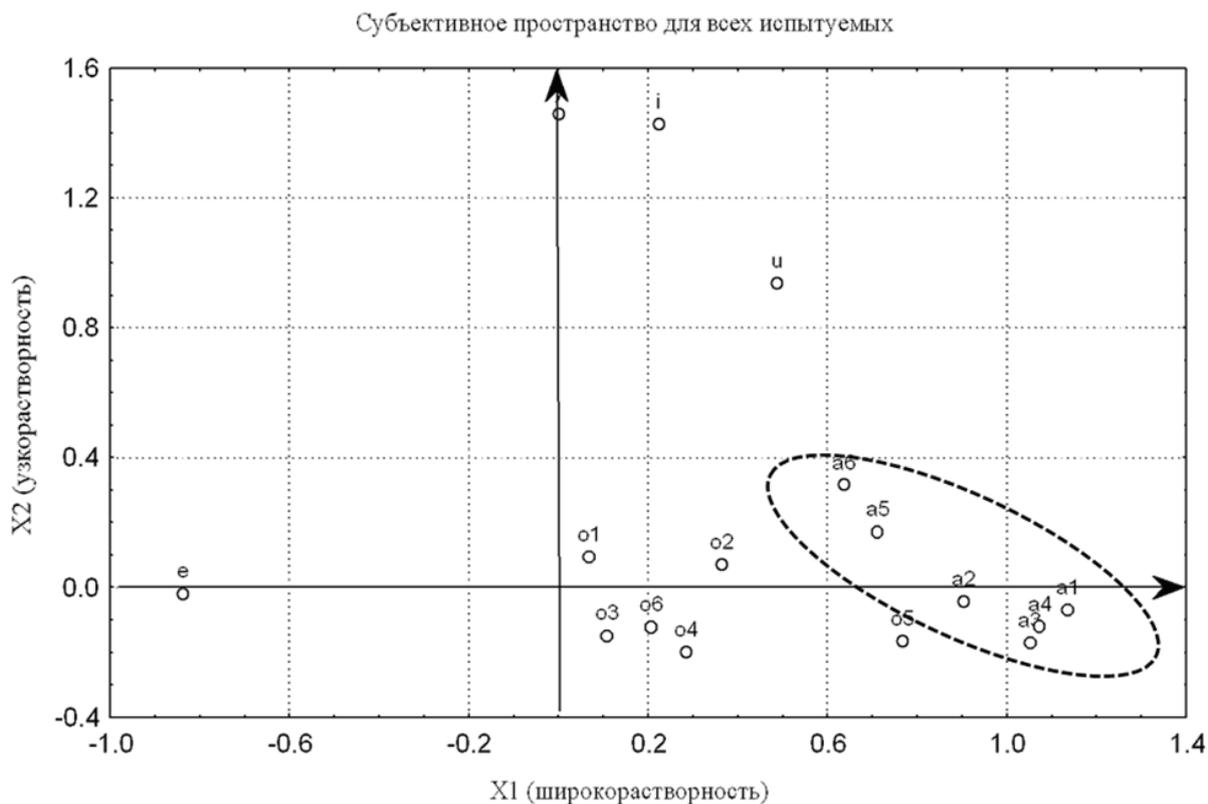


Рис. 4. Субъективное пространство восприятия гласных звуков русского языка в проекции на плоскость осей 1 и 2. Точки подписаны условными транскрипциями (латинские буквы). Кругом выделены варианты гласного «А», произнесенных с разной степенью широты раскрытия рта.

На рис. 4 видно, что разные варианты гласного «А», произнесенные с разной степенью широты раскрытия рта, располагаются вдоль оси 1, обозначенной как широкостворность.

На рис. 5 показана линейная зависимость воспринимаемого качества звучащего образца гласного «А» (ось 1 субъективного пространства) от степени широкостворности рта говорящего, что подтверждает гипотезу о том, что при восприятии гласных звуков происходит декодирование соответствующего артикуляционного паттерна говорящего, заданного соответствующей кинаемой широкостворности.

На рис. 6 представлено субъективное пространство восприятия в проекции на ось 1 и 3. На этом рисунке видно, что разные варианты гласного «О», произнесенные с разной степенью огубленности, располагаются вдоль оси 3, обозначенной как огубленность.

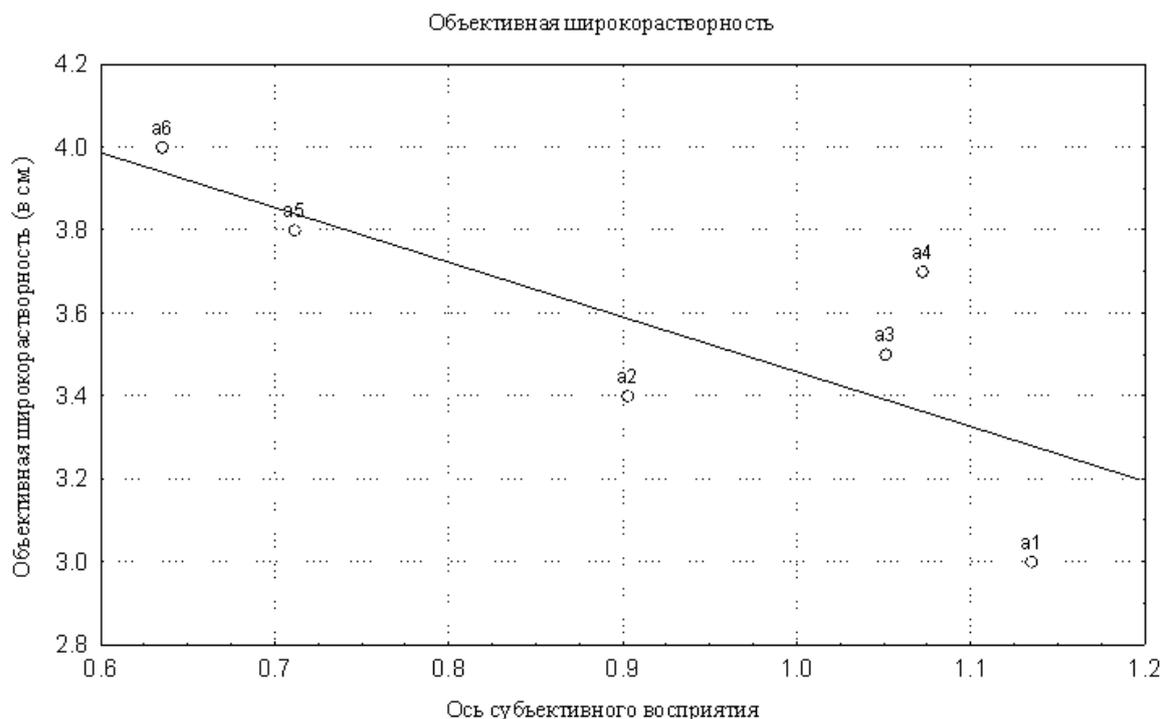


Рис. 5. Зависимость воспринимаемого субъективного качества звучащего образца гласного «А» (ось 1 субъективного пространства) и объективно (в см.) измеренной степенью широкораспорности при произнесении соответствующего звука.

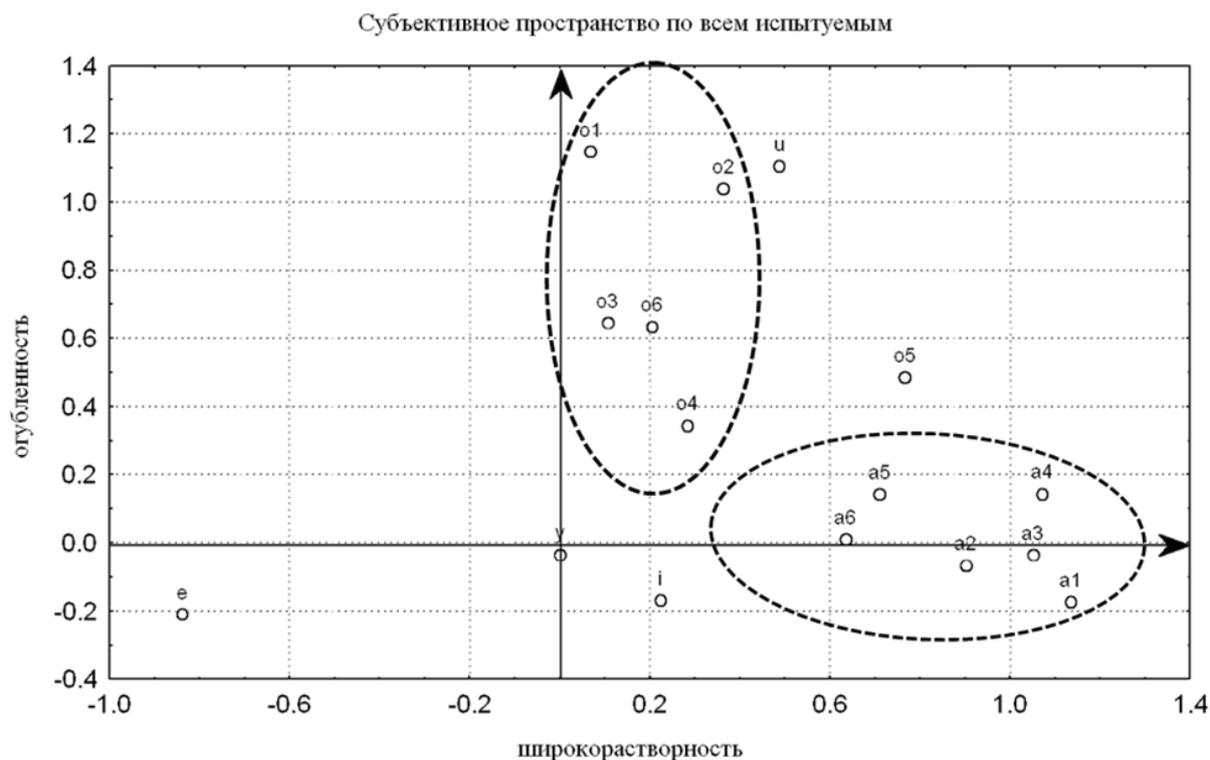


Рис. 6. Субъективное пространство восприятия гласных звуков русского языка в проекции на плоскость осей 1 и 3. Точки подписаны условными транскрипционными знаками (латинские буквы). Кругом выделены варианты гласного «О», произнесенные с разной степенью огубленности и гласного «А», произнесенных с разной степенью широты раскрытия рта.

На рис. 7 показана линейная зависимость воспринимаемого качества звучащего образца гласного «О» (ось 3 субъективного пространства) от степени огубленности рта говорящего,

что подтверждает гипотезу о том, что при восприятии гласных звуков происходит декодирование соответствующего артикуляционного паттерна говорящего, заданного соответствующей кинаемой огубленности.

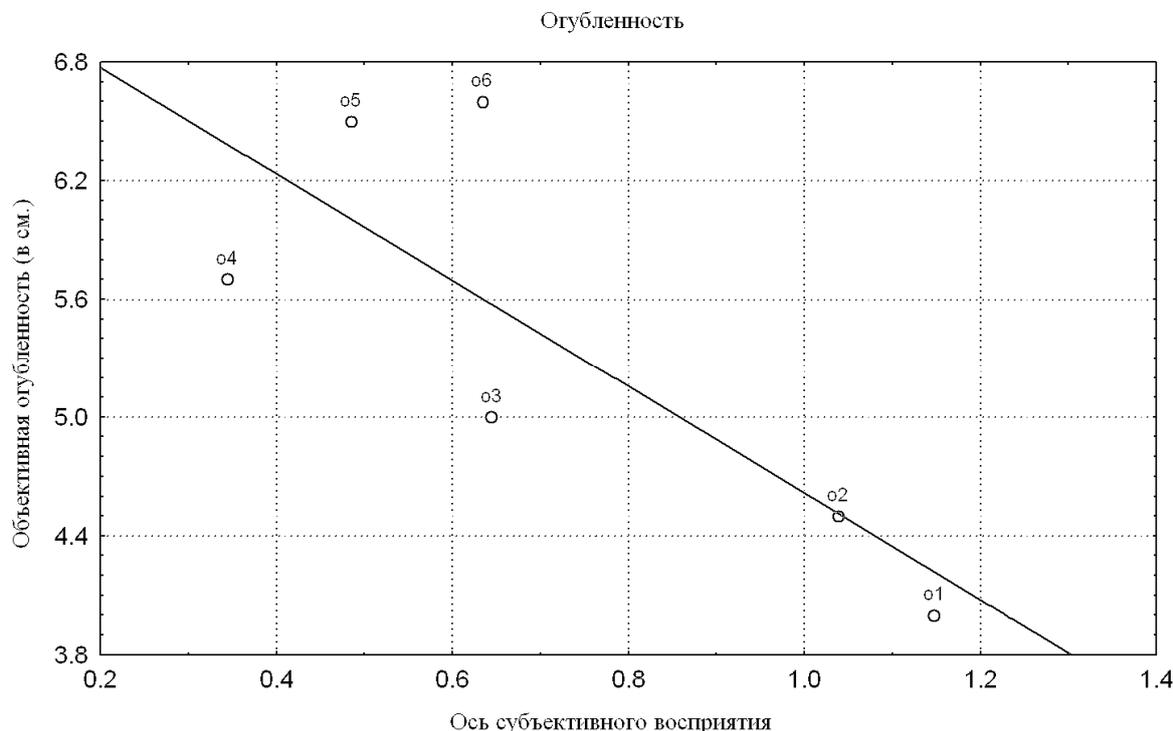


Рис. 7. Зависимость воспринимаемого субъективного качества звучащего образца гласного «О» (ось 3 субъективного пространства) и объективно (в см.) измеренной степенью огубленности при произнесении соответствующего звука.

Зарегистрированные при восприятии гласных звуков ВП мозга (в 16 отведениях) были обработаны методом факторного анализа, и было установлено, что форма усредненного ВП совпадает со стандартизированной формой фактора, отражающего в своей амплитуде динамику изменений ВП для разных вариантов широкорастворности звуков «А». Иными словами, форма (временной паттерн и мозговая локализация) вызванных ответов одинаковы при разных вариантах произношения звука «А», кодирующего одну и ту же кинаему широкорастворности, а различия в степени выраженности данного артикуляционного признака проявляются только в амплитуде ВП. В результате локализации мозговой активности выявлено совместное участие в процессе восприятия и анализа отдельных гласных звуков «А» таких структур как левая височная зона коры (Вернике) и бледный шар. На рис. 8 для примера показан ВП (в отведении F4) для одного испытуемого и результаты динамической локализации мозговой активности.

На рис. 9 показана факторная нагрузка стимула (выраженности амплитуды ВП) от субъективного восприятия звука «А» (ось 1 субъективного пространства), произносимого с разной степенью широкорастворности. Видна прямая зависимость этих показателей, из чего следует, что амплитуда ВП при данной инструкции отражает количественные различия стимулов.

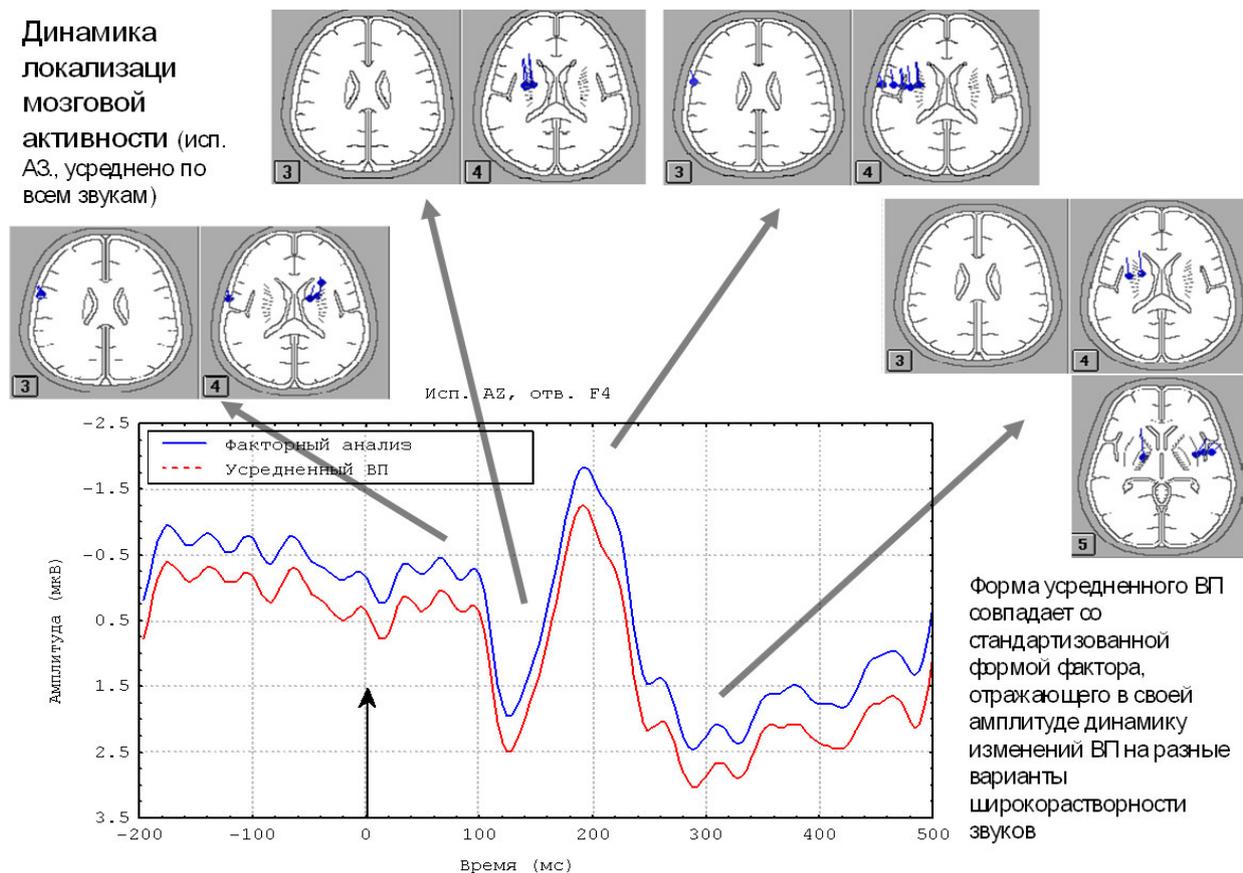


Рис. 8. Пример вызванного усредненного ответа мозга (в отведении F4, пунктирная линия) и факторного представления (сплошная линия) для одного испытуемого, а также основные этапы локализации активности мозговых структур, выявляемые с помощью дипольного анализа. На графике момент предъявления звука (значение 0 по шкале времени в мс) отмечен вертикальной стрелкой. Стрелки к фрагментам атласа мозга с выявленными кластерами диполей показывают соответствие по времени компонентов ВП и локализацией мозговой активности.

Таким образом, на основании психофизиологических экспериментов с восприятием гласных звуков русского языка подтвердилась гипотеза об артикуляционной (кинакемной) природе единиц восприятия речи. Полученные оси субъективного пространства удалось интерпретировать как кинакемные параметры широкорастворности, узкорастворности и огубленности.

В третьей серии экспериментов исследовалось восприятие согласных звуков русского языка. В соответствии со схемой многомерного шкалирования, набор из 23 согласных звуков, произнесенных в слоге с последующим гласным «и» или «ы», предъявлялся 6 испытуемым для оценки степени воспринимаемого различия. Всего было 253 пары стимулов, предъявляемых в случайном порядке по несколько раз (по всем предъявлениям для всех испытуемых было 10 реализаций каждой пары). Результаты оценок сведены в матрицу различий и построена 8-мерная пространственная модель, которая с коэффициентом корреляции в .96935 воспроизводила систему экспериментальных оценок. Полученное распределение исследуемых звуков представлено на рис. 10-11 проекциями на соответствующие пары осей.

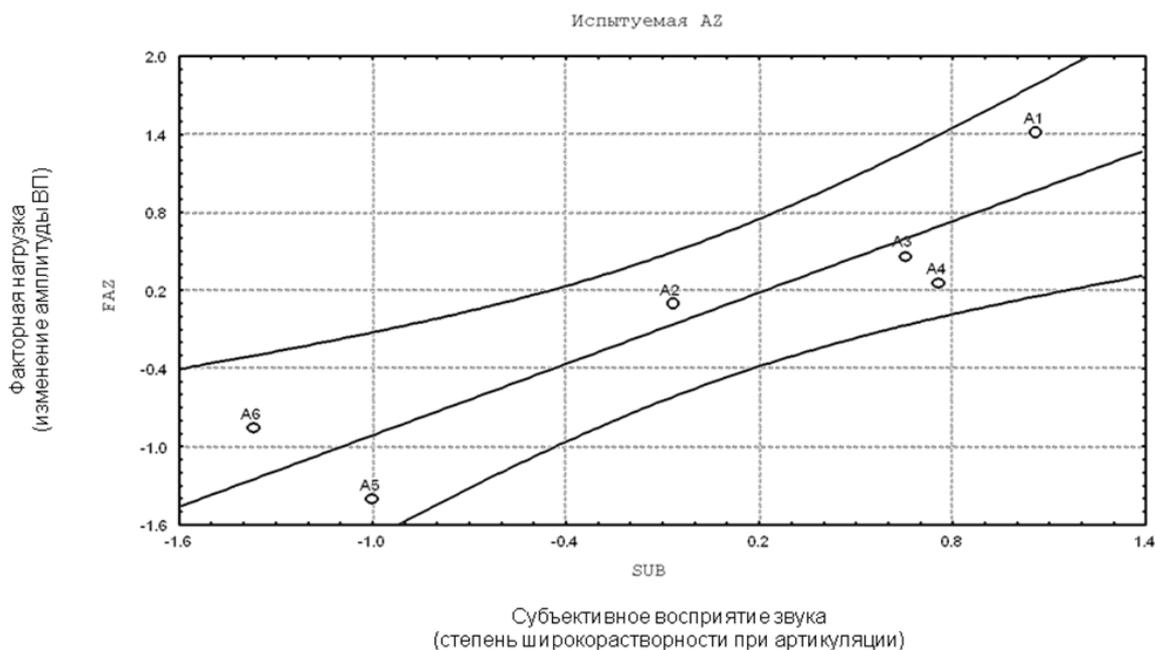


Рис. 9. Зависимость амплитуды ВП (по факторной нагрузке) от субъективного восприятия (значения по оси 1 субъективного пространства) а так же, как было показано ранее, от ширины раскрытия рта говорящего.

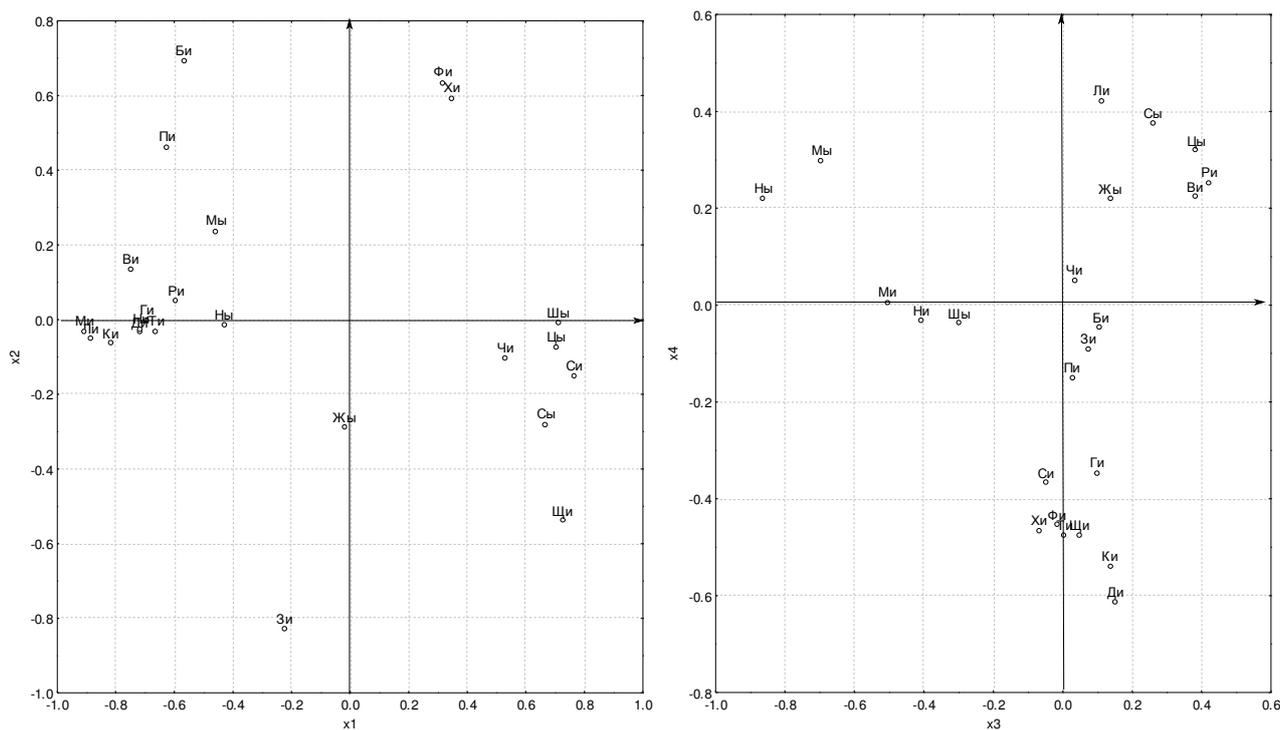


Рис. 10. Проекция на оси 1 - 2 и 3 - 4 пространства восприятия согласных.

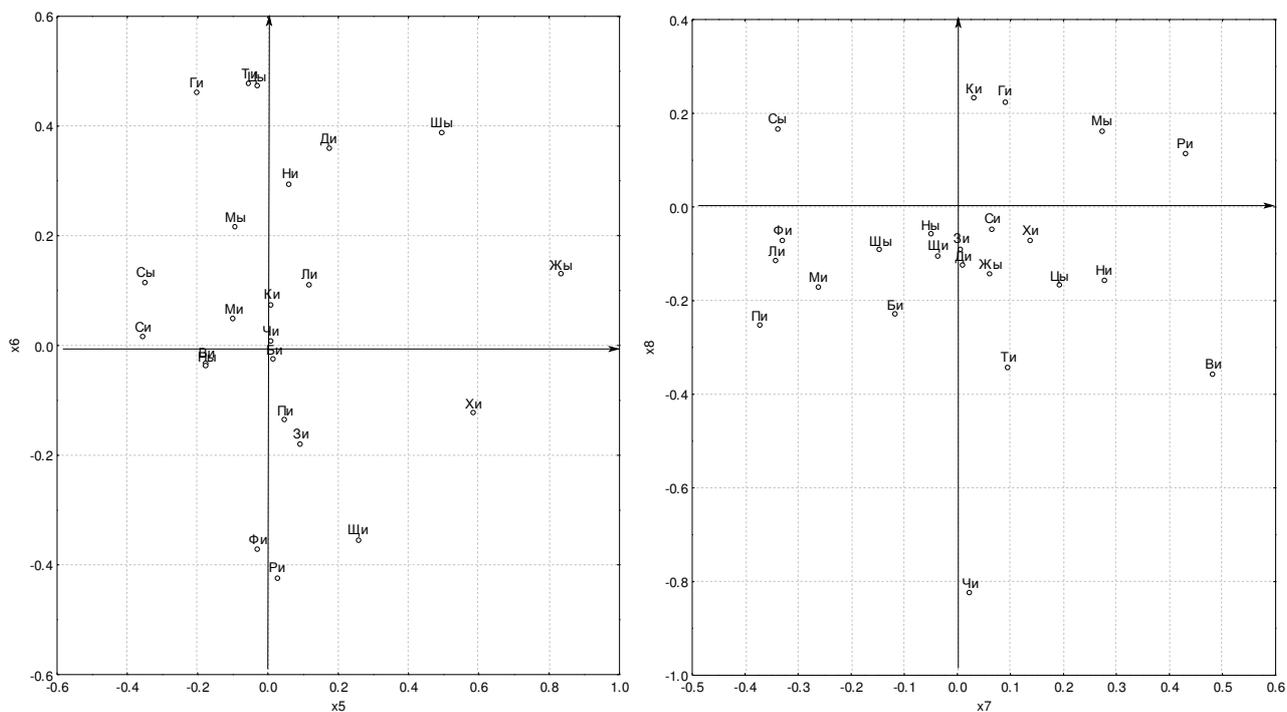


Рис. 11. Проекция на оси 5 - 6 и 7 - 8 пространства восприятия согласных.

Полученные оси пространства восприятия согласных звуков можно интерпретировать, по аналогии с осями для гласных, как соответствующие кинакემные оппозиции, в данном случае консонантной подсистемы, в соответствии с таблицей В.Я. Плоткина для русского языка (см. рис. 1).

Таким образом, на основании всех психофизиологических экспериментов гипотеза об артикуляционной (кинакемной) природе единиц восприятия речи подтвердилась.

Литература

1. Бодуэн де Куртене И.А. Избранные труды по общему языкознанию. М., 1963. Т.1, 2.
2. Варганов А.В. Многофакторный метод разделения ЭЭГ на корковую и глубинную составляющие. Журн. высш. нервн. деят. 2002. т.52. N 1. 111-118.
3. Варганов А.В., Варганова И.И. Онтогенез восприятия и категоризации эмоций. Вестник РГНФ, N 3, 2004, с. 197-207.
4. Варганов А.В., Варганова И.И. Эмоции, мотивация, потребность в филогенезе психики и мозга. Вест. Моск. Ун-та. Сер. 14 Психология. 2005. N3 с.20-35.
5. Варганов А.В., Виденеева Н.М. Четырехмерная сферическая модель эмоций и дистанционный речевой контроль состояния человека. Тезисы докладов рабочей группы «Влияние информационных технологий на национальную безопасность» 4-й Ежегодной Конференции Консорциума ПрМ «Построение стратегического сообщества через образование и науку». Москва, 25-27 июня 2001 г., 35 с.
6. Виденеева Н.М., Хлудова О.О., Варганов А.В. Эмоциональные характеристики звучащего слова. Журнал ВВД. 2000. т.50 вып. 1, с. 29-43.
7. Плоткин В.Я. Фонологические кванты. – Новосибирск: ВО «Наука». 1993, 180 с.
8. Щерба Л.В. Языковая система и речевая деятельность. Л., 1974.
9. Якобсон Р.О. Избранные работы. М., 1985.

Ал.Б. Леонтьева,¹ Р. Йекель,² И.А. Кагиров,¹ А.Л. Ронжин¹

Анализ речи и поведения студентов при компьютерном тестировании знаний в многомодальном режиме

*(1) Санкт-Петербургский институт информатики и автоматизации РАН,
г. Санкт-Петербург, Россия*

leonty@ias.spb.su

*(2) Дрезденский технологический университет,
г. Дрезден, Германия*

Rainer.Jaeckel@ias.et.tu-dresden.de

Введение

При разработке диалоговой системы необходимо учитывать предпочтения потенциальных пользователей, относительно способа общения и типа интерфейса. Сравнивая особенности многомодального (для ввода информации допускается использование речи и сенсорного экрана) и унимодального (или речь, или сенсорный экран) взаимодействий, можно сделать некоторые предположения о характере поведения и способах ввода информации и стиля речи при взаимодействии пользователя с машиной.

Большинство вероятных пользователей уже обладают некоторым опытом общения с компьютерными системами, потому что в повседневной жизни им приходится сталкиваться с получившими широкое распространение «информационными киосками» — обслуживающими автоматами, использующими стандартные средства ввода (клавиатура или набор функциональных кнопок) в комбинации с сенсорным экраном. Поэтому есть вероятность, что и при взаимодействии с многомодальной системой они будут придерживаться алгоритма взаимодействия, выработанного до этого. Например, человек, привыкший к тактильному характеру взаимодействия с компьютерной системой, даже при добавлении речевой модальности ввода будет охотнее пользоваться тактильной системой ввода, чем речевым запросом.

Во время традиционной межчеловеческой коммуникации люди почти всегда взаимодействуют многомодально. Естественно, при этом число источников информации или модальностей, которые доступны собеседнику, ограничено. Также и многомодальные системы ограничены в числе и типе входных модальностей, которые они могут распознавать. Кроме того, пользователи могут комбинировать виды ввода во время человеко-машинного взаимодействия, общаться многомодально или использовать только одну модальность. При этом, хотя пользователи и предпочитают взаимодействовать многомодально, а не одномодально, это абсолютно не гарантирует, что они будут подавать каждую команду системе многомодально [1]. Поэтому при разработке системы необходимо решить: пользователи будут работать одномодально или многомодально.

Исследования показывают что, при работе пользователей с визуально-пространственными объектами примерно в 20% случаев управление осуществляется многомодально, а остальное время отдаются только речевые команды или символы, сделанные световым пером [2]. Предсказать, будет ли человек выражать команды многомодально можно, зная тип действия, которое он выполняет в данный момент. В частности, пользователи почти всегда выражают команды многомодально, когда описывают пространственную информацию о расположении, размере, ориентации, форме объекта.

В настоящей статье будут рассмотрены особенности взаимодействия пользователя с диалоговой системой. Будут рассмотрены проблемы, связанные с факторами, влияющими на общение человека с компьютерной системой, и проанализирована необходимость предварительных знаний о поведении пользователя. В конце статьи будет рассказано об одном эксперименте, проведенном для изучения речи и поведения пользователей при взаимодействии с многомодальным киоском.

1. Особенности взаимодействия пользователя с диалоговой системой

На поведение человека в разной степени оказывают влияние следующие факторы: (1) контекст и особенности задачи, решаемой посредством диалоговой системы; (2) наличие и особенности предшествующего опыта, связанного с диалоговым взаимодействием (опыт может быть как персональным, так и опосредованным, может быть положительным (удачным) и отрицательным (неудачным)); (3) влияние интерфейса диалоговой системы на поведение пользователя, влияющего на выбор пользователем определенной модальности [3]. Ориентация на предшествующий опыт усиливается стереотипностью мышления, и при построении аналогий новый киоск попадает в одну категорию со всеми остальными автоматами, и, следовательно, стимул к новому алгоритму действий невелик. Однако при визуальной и смысловой новизне объекта взаимодействия влияние опыта может уменьшиться, так как новый объект не будет идентифицироваться ни в одну из уже существующих мысленных категорий, а, значит, будет требовать и нового подхода при работе с ним. Положительный предшествующий опыт, будь он персональным или опосредованным (кто-то сказал, где-то увидел и т.д.) будет склонять решение в пользу повторения опыта, в то время как отрицательный опыт способствует поиску альтернатив, делая более вероятной возможность выбора альтернативной модальности. Важно также влияние самой диалоговой системы, и ее интерфейса в особенности, на выбор и предпочтения пользователя. Вид интерфейса и способов реализации взаимодействия, безусловно, могут в значительной мере определять поведение пользователя с системой. Например, четко прорисованные на сенсорном экране виртуальные кнопки неявно привлекают внимание и способствуют выбору именно тактильного (сенсорного) взаимодействия, в то время как простой список из пунктов меню не привлекает внимания пользователя столь же сильно, что заставляет сомневаться в возможности тактильного ввода и искать другие способы взаимодействия.

Другой аспект рассматриваемого вопроса о взаимодействии пользователя с системой заключается в непреднамеренном создании стресса у неподготовленного пользователя при неоднозначности интерфейса [4,2]. Непонимание, какую модальность использовать для управления приводит к стрессу, а значит, должен быть тщательно продуман интерфейс. Если в интерфейсе нет очевидного указания на ожидаемый от пользователя характер взаимодействия с системой – пользователь теряется, не зная, каким образом строить свое поведение. Если в системе не предусмотрено наглядных для пользователя реакций на его действия – опять создается ситуация стресса из-за неуверенности пользователя в правильном характере взаимодействия с системой. Для избежания таких ситуаций помимо проработки интерфейса, с учетом особенностей потенциального пользователя, возможно введение «обратных связей». Это могут быть как подтверждения принятия ответа, так и варианты ответов, предлагаемые пользователю, запрос отзыва от пользователя по окончании взаимодействия, четкое обозначение каждого действия системы. Все эти «способы общения» скажутся на естественности взаимодействия, поставив пользователя в четкие рамки и подтверждая и указывая/оговаривая порой очевидные для данного конкретного пользователя вещи.

2. Знания о поведении пользователя, используемые при построении многомодальной системы

Основной задачей данного исследования являлось оценивание поведения пользователя и его предпочтений при работе с прототипом многомодального киоска, а также анализ особенностей спонтанных речевых ответов информантов. Априорно перед началом эксперимента были сделаны некоторые выводы о поведении пользователя в зависимости от способа ввода информации и типа предлагаемого интерфейса.

При визуальном оформлении, дающем всю полноту информации, когда на экране выводится вопрос, предлагаются варианты ответов, которые пользователь может выбрать, нажав соответствующие области («кнопки») на сенсорном экране, можно ожидать минимум речи от пользователя, особенно относящейся к киоску. Потенциальный клиент видит интерфейс, предполагающий тактильно-визуальное взаимодействие, и может не рассматривать возмож-

ность обработки речевой модальности, даже если эта функция предусмотрена (вывод на основе имеющегося опыта), отдавая предпочтение указательным движениям на сенсорном экране, а не речевому взаимодействию. Однако предрасположенность к выбору сенсорного взаимодействия вместо речевого будет зависеть также и от визуальной составляющей интерфейса (насколько области с вариантами ответов будут напоминать «кнопки»). Вполне возможно присутствие заполненных и незаполненных пауз, вызванных размышлением.

При визуально-речевом варианте интерфейса, когда вопрос проговаривается речью без дублирования текстом на экране, а варианты ответов выводятся на мониторе, присутствие речевого обращения может привести к увеличению использования речи, в том числе и обращенной к киоску (элемент провоцирования на разговор) по сравнению с полностью визуальным интерфейсом. Возможно увеличение вокализации пауз в речи пользователя.

При полностью речевом (ориентированном на речь) интерфейсе (даже при наличии визуальной составляющей в системе), когда речью задается и вопрос, и предлагаемые варианты ответов, специфика такого диалога может привести к предпочтению речевого интерфейса, даже при наличии возможности сенсорного взаимодействия, но не столь очевидной как «кнопки» с вариантами ответа. Выбор речевого интерфейса может быть осуществлен как единственного кажущегося или как действительно присутствующего интерфейса. Естественно ожидать появления различного рода заминок, «небеглостей» в речи пользователя, например, переспрашивания вопроса. Скорее всего, будут присутствовать большие паузы, вызванные размышлением, особенно связанным с непривычностью к речевому интерфейсу, незнанием (и выбором) как лучше сформулировать ответ для речевой системы для успешного взаимодействия.

При полностью речевом интерфейсе (даже при наличии визуальной составляющей в системе), когда речью задается вопрос без предложенных вариантов ответа, по сравнению с описанным ранее речевым интерфейсом возможно появление спонтанной речи, спонтанно сформулированных фраз. Большие паузы будут вызваны размышлением над построением ответа/запроса, больше связанным с отсутствием предложенных шаблонов ответа.

При речевом и визуальном интерфейсе с полным дублированием речи текстом можно предположить поведение пользователя аналогичным варианту с полностью визуальным интерфейсом, вследствие большей информативности визуальной составляющей.

Все рассмотренные случаи, когда выбор отдается тому или иному интерфейсу, возможны только для тех задач диалогового взаимодействия, в которых нет явной потребности в совмещении интерфейсов (например, речевого и сенсорного), в отличие, например, от задач, связанных с многомодальным взаимодействием для разработки строительных чертежей [3]. В зависимости от типа задачи и типа информации, с которой происходит работа, а также от предшествующего опыта, пользователь для ввода информации будет выбирать необходимую модальность (в случае возможности выбора), или склоняться к выбору какой-то одной модальности, в рамках которой проведет весь диалог, или использовать несколько модальностей одновременно. Исходя из тех же факторов, влияющих на выбор модальности для ввода информации, следует строить интерфейс вывода. В данной статье скорее затрагивается изучение модальностей ввода, так как модальность вывода для пользователя играла пассивную роль получения априорного знания.

3. Описание эксперимента

Для проведения экспериментов использовался разработанный ранее в лаборатории исследовательский стенд с многомодальным пользовательским интерфейсом («многомодальный информационный киоск») [8]. Стенд объединяет стандартные средства ввода/вывода информации (сенсорный экран, клавиатуру и стерео динамики) и устройства, обеспечивающие бесконтактное взаимодействие (видеокамера, микрофоны). Для дистанционного распознавания русскоязычных голосовых команд применяется массив микрофонов (до 8 микрофонов), позволяющий локализовать источник полезного сигнала и понизить влияние акустических шумов. Миниатюрная видеокамера обеспечивает оптическое отслеживание положе-

ния пользователя в рабочей зоне киоска. Разработанный прототип киоска используется для создания информационно-справочных систем и исследования когнитивных аспектов взаимодействия пользователя с системой, учитывая различные способы коммуникации.

Эксперименты проводились для трех вариантов интерфейса: (1) полностью речевой интерфейс; (2) интерфейс с дублированием речи текстом; (3) полностью текстовый интерфейс. Тестирование было осуществлено посредством задавания пользователю ряда вопросов через определенный вид интерфейса. Было составлено два вида вопросов: с предложенными вариантами ответов и без. Тестирование трех вариантов интерфейса осуществлялось на вопросах первой категории. Для получения спонтанного ответа от пользователя и оценки его поведения в рамках спонтанного взаимодействия использовался второй вид вопросов, заданных через текстово-речевой интерфейс. Для определения влияния опыта на последующее взаимодействие для разных групп пользователей было сделано чередование последовательности групп вопросов первого и второго видов.

Всего были опрошены 30 студентов технических специальностей с последних курсов ГУАП. Каждому информанту были предложены 20 вопросов, из которых первые десять имели четыре варианта ответа, а на оставшиеся десять требовалось ответить, дав собственное определение. Вопросы с вариантами ответа предлагались в трех разных режимах, в соответствии с используемым видом интерфейса. Режим (1) подразумевал вопрос, заданный синтезированным голосом, режим (2) — вопрос, заданный синтезированным голосом и продублированный текстом на экране и, наконец, режим (3) — это только текст. Опрашиваемые студенты были разделены на три группы по 10 человек, каждая группа опрашивалась только в одном из трех режимов. Группы делились на подгруппы по 5 человек, для которых различалась последовательность вопросов первого и второго видов. Для одной подгруппы вначале шли вопросы с вариантами ответов, для другой, наоборот, — без вариантов ответов. На вопрос первой категории давалось время 15 сек., на вопрос второй категории — 30 сек. На экран выводился индикатор времени («*progress bar*»), и, кроме того, на экране присутствовало сообщение о том, что речь информанта записывается. Запись происходила постоянно во всех трех режимах, даже когда респондент отдавал предпочтение только тактильному вводу, например, в режиме 3.

Тестирование проводилось с использованием методики «*Wizard of Oz*», обеспечивающей ручную корректировку результатов системы распознавания речи и управление процессом тестирования [5]. Данная методика направлена на оценивание пользовательских интерфейсов, изучение естественно-языковых интерфейсов и особенностей употребления естественного языка в информационных системах, моделирование человеко-машинного взаимодействия [6,7]. С помощью методики «*Wizard of Oz*» человек-оператор контролирует процесс работы системы и вручную выполняет часть или все операции системы скрыто от пользователя, взаимодействующего с системой. Пользователь, не зная о действиях оператора по эмуляции некоторых функций системы, полагает, что взаимодействует с реальной полностью функционирующей компьютерной системой, работающей в автоматическом режиме. Таким образом, удобно осуществлять наблюдение за пользователем, контролируя процесс взаимодействия, давая пользователю полную свободу действий — диалог всегда будет успешно завершен. У оператора, следящего за диалогом, была возможность вывести на экран следующий вопрос, не дожидаясь окончания времени, отпущенного на предыдущий. В данном эксперименте система не анализировала ответ пользователя, а методика «*Wizard of Oz*» применялась исключительно в целях сокращения общего времени тестирования.

Контакт информантов с киоском происходил в обычном (т.е. не звукоизолированном) помещении, в присутствии оператора. Таким образом, тестируемый студент попадал в ситуацию взаимодействия с многомодальной системой, размещенной в общественном месте. Респонденту не давалось специальных указаний относительно того, должен ли он реагировать на присутствие других людей, то есть он вполне мог обратиться за помощью к окружающим. Некоторые из студентов искали помощи у тестирующих, но помощи не получали, поэтому дальше старались взаимодействовать исключительно с киоском. Тестирование сту-

дентов производилось по одному — последующие студенты не видели ход проведения предыдущего эксперимента.

4. Результаты экспериментов

При проведении экспериментов велась постоянная запись ответов студентов и мониторинг нажатия кнопок на сенсорном экране. В данном разделе проводится анализ реакций студентов по накопленным аудиозаписям и файлу, содержащему системную информацию о работе киоска и действиях пользователей (так называемому «лог-файлу»). В таблице 1 приводятся типы и количество явлений. Под явлениями понимаются реакции информантов, зафиксированные во время проведения эксперимента.

Таблица 1. Реакции, зафиксированные в ходе эксперимента

	Вопрос к окружающим	Попытка управлять диалогом с киоском	Молчание	Озвученная пауза	Размышления вслух	Самоисправления	Множественное нажатие на кнопку	Повторение ответа
Количество явлений	24	6	37	41	82	17	3	5
Количество студентов	15	4	15	15	22	8	3	3

Половина опрашиваемых студентов искали помощи у людей, а не у киоска — возможно, не доверяя «машине» или не видя каких-то функциональных возможностей для того, чтобы обратиться к киоску. Это было характерно для тех моментов, когда пользователь или не был уверен в том, какую входную модальность использовать («А здесь *кнопочку* нажимать?»), не был уверен в характере требуемого ответа («А тут как, *просто номер называть?*»), или попросту не знал, как вести себя в ситуации, когда не знаешь ответа («А *что говорить*, если не знаешь ответа?»).

Большинство тестируемых студентов, испытывая затруднения с ответом, предпочитали молчать, дожидаясь следующего вопроса. Иногда эта пауза вокализовалась растягиванием гласных или согласных, громким выдохом, смехом, прочищением горла, кашлем. И наоборот, если ответ на вопрос был известен респонденту, то ответ давался незамедлительно и, в случае речевого взаимодействия, проговаривался быстрее, даже несмотря на наличие запаса времени для ответа на вопрос. Часто процесс раздумий и ответа на вопрос сопровождался так называемым размышлением вслух: чтением вслух вопроса (в том числе с вариантами ответов), размышлением вслух в процессе нахождения ответа, нажатием кнопки на сенсорном экране с одновременным проговариванием ответа, высказыванием своего мнения о работе системы и своих знаниях, дискурсивными оборотами, образующими связки в диалоге между людьми, вопросами, не требующими ответа и др. Иногда в процессе такого размышления и подбора ответа вслух пользователь сам поправлял себя, оговаривался, торопясь произнести правильный ответ. Эти явления представлены в таблице 1 как самоисправления.

Для ряда студентов была характерна некоторая поспешность действий, выражающаяся в том, что они многократно нажимали на сенсорную кнопку и повторяли ответ. Следует особо отметить, что диалог был «односторонним», т.е. система только получала информацию от пользователя и выводила информационные сообщения о принятии ответа. Перед тестируемым не стояло задачи получения чего-либо от системы, и значит, не было страха перед неудачным взаимодействием, например, вследствие неверно распознанной фразы, чего некото-

рые могли бы попытаться избежать, повторив ответ. Последующий опрос студентов показал, что некоторых из них смущало присутствие столь длительного времени на ответ, особенно, когда они тратили на ответ гораздо меньшее время («Может какие-нибудь *длинные определения надо давать?*»): как правило, вначале они отвечали односложно, а потом давали длинное определение, стремясь использовать все оставшееся время.

Таблица 2. Статистическая характеристика результатов экспериментов.

Структура тестов				Полученные ответы	Молчание	Озвученная пауза	Сенсор и молчание	Сенсор и вокализация	Попытка управления диалогом вместо ответа / с ответом	Вопрос к окружающим вместо ответа / с ответом
Номер теста	Порядковые номера вопросов	Вид вопросов	Тип интерфейса							
1a	1-10	с ответами	речь	48	0	2	-	-	0 / 1	0 / 0
	11-20	без ответов	текст-речь	39	5	4	-	-	1 / 1	1 / 2
1b	1-10	без ответов	текст-речь	44	3	3	-	-	0 / 0	0 / 5
	11-20	с ответами	речь	49	0	1	-	-	0 / 2	0 / 4
2a	1-10	с ответами	текст	0	0	0	37	12	0 / 0	1 / 0
	11-20	без ответов	текст-речь	44	2	3	-	-	0 / 0	1 / 1
2b	1-10	без ответов	текст-речь	37	6	5	-	-	0 / 0	2 / 3
	11-20	с ответами	текст	10	0	0	29	10	0 / 0	1 / 0
3a	1-10	с ответами	текст-речь	19	0	1	27	3	0 / 0	0 / 2
	11-20	без ответов	текст-речь	27	10	12	-	-	1 / 0	0 / 0
3b	1-10	без ответов	текст-речь	35	5	9	-	-	0 / 0	1 / 0
	11-20	с ответами	текст-речь	18	6	1	25	0	0 / 0	0 / 0

Таблица 2 показывает результаты, разбитые по видам экспериментов. Тесты пронумерованы от одного до трех, обозначая разные используемые режимы вывода информации для первой категории вопросов: речь, текст, текст с речью. Индексы «а» и «b» обозначают группы с разной последовательностью вопросов первой и второй категории.

Под ответом пользователя подразумевается ответ на вопрос любым образом: речью или нажатием кнопки; для вопросов с вариантами возможно или повторение варианта ответа или только его номера; или допускается обозначение отсутствия знаний по данному вопросу, выраженное, например, «не знаю». В отличие от реального экзамена, здесь студенты, не зная ответа, признавали это, не пытаясь что-то придумать, или молчали, ожидая следующего вопроса. В вопросах с предложенными вариантами ответов свою неуверенность в выбранном варианте обозначали словами «вроде», «пусть будет», «это что-то вроде» и др. Для ответов тестируемых, не смотря на общение с машиной, а не с человеком, была характерна более размытая формулировка. Большинство ответов не были похожи на краткие команды, а дополнялись другими словами, характеризующими ответ, степень уверенности в нем, размышлениями вслух и др. Также для ответов пользователей (особенно при речевом интерфейсе)

были свойственны переспрашивания, если плохо запомнился вопрос или варианты ответов. При этом респонденты не ждали ответа на эти вопросы, а использовали их только как дискурсивные маркеры, часто употребляющиеся в диалоге между людьми.

При прохождении тестов 1a и 1b многие студенты, предполагая, что последующие вопросы также будут задаваться только посредством речи, стояли вполборота к киоску, не смотря на экран, а только слушая. В тестах 2 и 3 появляется возможность использования сенсорного монитора для ответа. В 2b один студент и по два студента в 3a и 3b тестах использовали только речевой интерфейс, большинство же предпочли давать ответы, нажимая виртуальные кнопки на сенсорном мониторе. При дальнейшем опросе было выяснено, что некоторые использовали сенсорное взаимодействие, т.к. видели курсор мыши на мониторе (присутствовал постоянно), а другие использовали речевое взаимодействие, сосредоточив внимание на сообщении о записи речи (присутствовало постоянно). Интерфейс, через который осуществлялось взаимодействие, не предрасполагал к какому-либо определенному характеру взаимодействия. Поэтому можно утверждать, что выбор тестируемых студентов опирался только на их собственный опыт и знания. Влияние предшествующего опыта оказалось несущественным, т.к. наблюдалось только три случая использования речевого интерфейса вместо сенсорного при предшествующем речевом взаимодействии и двум случаям выбора не предшествовал никакой другой интерфейс.

Интересно отметить, что молчание чаще всего фиксировалось после вопросов без вариантов ответов, тогда как вопросы с вариантами ответов, в основном, сопровождались вокализацией молчания. Это явление, безусловно, заслуживает более пристального рассмотрения и требует дополнительного анализа.

5. Заключение

Эксперимент показал, что потенциальный пользователь многомодального киоска не готов пользоваться всеми функциональными возможностями, предлагаемыми системой. Так, принявшие участие в эксперименте студенты предпочитали пользоваться сенсорным вводом, а не речевым; это не исключало, впрочем, использование обеих модальностей — некоторые респонденты в одинаковой степени использовали и речевой, и сенсорный ввод.

Кроме того, было зафиксировано несколько попыток управления диалогом (фразы типа: «Дальше!», «Можно следующий вопрос?»), что указывает на определенную степень доверия к компьютеру.

Предполагается, что ограничения на использование определенного вида интерфейса, которые пользователи сами накладывали на себя, вызваны специфичностью задачи, не требующей определенного характера взаимодействия, отсутствием опыта общения с подобными системами, а также чисто психологическими установками. Попытки выйти за рамки этих ограничений — использование всех модальностей, попытки контролировать диалог и т.п. — внушают уверенность, что со временем потенциальный пользователь будет использовать возможности киосков многомодального типа гораздо шире, чем сейчас.

Описанный в настоящей статье эксперимент, конечно же, не претендует на полную демонстрацию проблем человеко-машинного взаимодействия; исследования следует повторить на более широком материале. Однако мы надеемся, что данный эксперимент поможет всем, кто занимается исследованиями в области речевых и других типов интерфейсов, обеспечивающих естественное взаимодействие человека с компьютером.

Литература

1. Oviatt, S. L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction. Special issue on Multimodal Interfaces*, 12, 93-129.
2. А.Л. Ронжин, А.А. Карпов, И.В. Ли. Речевой и многомодальный интерфейсы. - М.: Наука, 2006 - (Информатика: неограниченные возможности и возможные ограничения), 173 с.

3. Bosch ten L., Boves L. Survey of spontaneous speech phenomena in a multimodal dialogue system and some implications for ASR // INTERSPEECH-2004, 1505-1508.
4. Stedmon, A. W. and Baber, C. Evaluating Stress in the Development of Speech Interface Technology. In Proceedings of HCI international (the 8th international Conference on Human-Computer interaction) on Human-Computer interaction: Ergonomics and User interfaces-Volume I. H. Bullinger and J. Ziegler, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, 1999, pp. 545-549.
5. Dahlback, N., Jonsson, A., Ahrenberg, L. Wizard of Oz Studies – Why and How. Knowledge-Based Systems, Vol. 6, No. 4, 1993. — P. 258-266.
6. А. Л. Ронжин, А. Б. Леонтьева. Применение методики «Гудвин» для моделирования человеко-машинного взаимодействия // Изв. вузов. Приборостроение. 2006. Т. 49, № 11. С. 70-75.
7. Salber, D., Coutaz, J. Applying the Wizard of Oz Technique to the Study of Multimodal Systems, EWHCI'93 Proceedings, East/West Human Computer Interaction, Moscow, August, 1993 (Lecture notes in Computer Science, Vol. 753, 1993. — P. 219-230).
8. А.Л. Ронжин, А.А. Карпов, Ан.Б. Леонтьева, Б.Е. Костюченко. Разработка многомодального информационного киоска. // Труды СПИИРАН. Вып. 5, т. 1. — СПб.: Наука, 2007, С. 227-245.

Научное издание

ВТОРОЙ МЕЖДИСЦИПЛИНАРНЫЙ СЕМИНАР
«АНАЛИЗ РАЗГОВОРНОЙ РУССКОЙ РЕЧИ» АР³ - 2008

Труды семинара

Составители: А.Л. Ронжин, И.А. Кагиров

Издание осуществлено за счет средств
гранта РФФИ № 08-07-06030-г.

Отпечатано в типографии «АНО Центр делового развития Василеостровский»
Санкт-Петербург, Большой пр. В.О., д. 55, тел./факс 321-87-02
Подписано в печать 04.07.2008 Печ. л. 5,5 тираж 200 экз., формат 291x207