Четвертый междисциплинарный семинар

«Анализ разговорной русской речи»

АР^3-2010

26–27 августа 2010 года,
Санкт-Петербург, СПИИРАН

Учреждение Российской академии наук Санкт-Петербургский
институт информатики и автоматизации РАН

Четвертый
междисциплинарный семинар

«Анализ разговорной русской речи»

$АР^3$ - 2010

26 – 27 августа 2010 года,
Санкт-Петербург, СПИИРАН

Санкт-Петербург
2010

УДК 004.522

Учреждение Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН

Санкт-Петербург, 199178, 14 линия, 39.
http://www.spiiras.nw.ru/speech

А64 Анализ разговорной русской речи» (АР^3-2010): Труды четвертого междисциплинарного семинара – СПб.: ГУАП, 2010.– 72 с.
ISBN 978-5-8088-0542-2

Издание представляет собой сборник докладов, сделанных на заседаниях четвертого междисциплинарного семинара «Анализ разговорной русской речи» (АР^3-2010), проходившего 26 – 27 августа 2010 года в Учреждении Российской академии наук Санкт-Петербургском институте информатики и автоматизации РАН. 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.

ISBN 978-5-8088-0542-2

© СПИИРАН, 2010
© Коллектив авторов, 2010
© ГУАП, оформление 2010

Оглавление

Л.А. Васильева, С.О. Тананайко. Исследование фонетической информации как основа для диагностики акцента и диалекта в многонациональном социуме…… 4

А.В. Венцов. Словесное ударение, собственная длительность гласных и ментальный лексикон……... 10

А.В. Венцов, Е.И. Риехакайнен. Восприятие омофонов и структура ментального лексикона……. 12

Н.А. Слепокурова, Е.П. Комовкина. Структура межпаузальных интервалов спонтанного диалога как фактор взаимодействия собеседников……. 16

М.О. Пономарь. Обработка речи на фонетическом уровне в интересах скрытного встраивания данных…….……..…… 23

С.И. Николенко, М.Л. Кореневский, И.А. Пономарёва, К.Е. Левин. Двухпроходное распознавание на основе классификации речи по тематике…….. 28

И.С. Кипяткова. Исследование статистических n-граммных моделей языка для распознавания слитной русской речи со сверхбольшим словарем……... 33

Ю.Г. Бондарос. Программно-аппаратный комплекс ”ТРЕНЕР” для тестирования систем распознавания речевых команд……. 41

М.В. Прищепа, В.Ю. Будков. Разработка интерактивной модели речевого взаимодействия мобильного информационного робота с посетителями торговых центров…….. 46

С.В. Глазков. Разработка интерактивных приложений с многомодальным интерфейсом для гетерогенных мобильных устройств…….. 51

М.А. Мясоедова, Д.Н. Душкин, З.П. Мясоедова, Н.В. Петухова, М.П. Фархадов. Разработка интернет портала «Сурдосервер» с ресурсами русского жестового языка.…… 57

С.В. Андреева. Лингвистические черты в зеркале систем автоматического понимания текста…… 62

Н.Б Вольская, С.Воробьёва. Нарушение линии деклинации как способ реализации акцентного выделения……... 64

Л.А. Васильева, С.О. Тананайко

Исследование фонетической информации как основа для диагностики акцента и диалекта в многонациональном социуме

Филологический факультет Санкт-Петербургского государственного университета, Санкт-Петербург, Россия
LAVasil@phonetics.pu.ru, stananaiko@mail.ru

1. Введение

Для проведения качественных лингвистических экспертиз, в частности, на фонетическом уровне, необходимо осуществление ряда фундаментальных и прикладных исследований. Теоретический аспект этих исследований должен быть направлен на изучение пределов допустимой вариативности фундаментальных свойств звуковой системы, так как результатом этой вариативности и являются региональные и акцентные варианты русского языка.

Прикладной аспект исследований включает решение следующих задач:

1) изучение состояния взаимодействия региолектов на территории современной России с учетом новой социальной и экономической ситуации в России;

2) системный анализ варьирования русского языка как неродного на фонологическом и фонетическом уровнях в внутри и за пределами России;

3) автоматизация поиска и интерпретации акустических характеристик сегментных и супрасегментных единиц регионально и акцентно окрашенной речи по сравнению с нормативной.

Исходная посылка данных теоретического и прикладного подходов состоит в том, что взаимодействие нормативного и ненормативного произношения будет давать специфический, допускающий идентификацию говорящего результат для каждого региона и родного языка говорящего. Предлагаемые подходы к осуществлению исследования – детальный акустический и аудиторский анализ фонетических характеристик спонтанной речи и чтение текста, реализованных дикторами мужчинами и женщинами, носителями региональных и акцентных вариантов и современной произносительной нормы русского языка. Желательно включение в исследование максимально широкого круга принадлежащих к различным группам носителей различных языков, а также региональных вариантов русского языка.

Результатом исследования станет методика распознавания, позволяющая идентифицировать родной язык говорящего (или один из диалектов этого языка), а также определить, какими еще языками он владеет.

Основным способом хранения и обработки звуковой информации была выбрана созданная на кафедре фонетики СПбГУ звуковая база данных для описания региональных и нормативных реализаций. База включает полученный материал и результаты его многоуровневой сегментации, акустической обработки, а также позволяет находить требуемые фрагменты звукозаписи и осуществлять поиск и систематизацию материала. К настоящему моменту эта базы данных заполнена образцами наиболее представительного звукового материала.

Необходимо добавить, что по нашему мнению, наряду с использованием при диагностике акцента различных математических и статистических методов очень важно получить подробные и надежные акустические данные по интерферированной речи дикторов с учетом данных предварительного слухового анализа на большом объеме качественных записей спонтанной речи. Надежность диагностики, несомненно, повышается при наличии такого исходного материала для анализа. Примером подобной детализированной диагностики является описание экспериментального исследования по выявлению

акустических коррелятов узбекско-русской интерференции при реализации элементов вокалической и консонантной русской фонетической системы (материал взят из вышеупомянутой БД).

Узбекский язык, включающий большое количество диалектов, которые имеют как много общего, так и специфического, характерного для каждого диалекта в отдельности, относится к группе тюркских языков, но при этом испытывает сильное ираноязычное влияние (диалектов таджикского языка).

Все эксперименты в настоящем исследовании проводились на основе записей русской слитной речи узбеков. Дикторы были в возрасте 20-40 лет, и их уровень владения русским языком при экспертной оценке был определен как средний. Все дикторы родились в области Навои, следовательно, они являются носителями ташкентско-ферганского «окающего» несингармонического говора. Материалом для анализа стал фонетически представительный текст (в дальнейшем - ФПТ) [1] в чтении четырёх дикторов и спонтанные фразы на русском языке, произнесённые теми же дикторами. Общая длительность звучащей речи составляла около 10 мин. для каждого диктора.

2. Вокализм

По результатам слухового анализа и подробной экспертной сегментации осуществлялся акустический анализ материала. Для выявления основных закономерностей была выполнена усредняющая статистическая обработка значений частот формант ударных и безударных гласных в произношении дикторов-узбеков в чтении ФПТ и в спонтанной речи. Анализ ограничен двумя формантами, как наиболее энергетически насыщенными, для них произведена сравнительная оценка отклонения частот от среднего значения.

Данные частотных значений, полученные в ходе исследования, сопоставлялись с данными значений для русской произносительной нормы, установленных в ходе новейших исследований [2].

Для акустического исследования были выбраны гласные /ɨ/, /a/, /o/, /e/, как наиболее подверженные изменению в условиях интерференции по результатам прогноза фонетических ошибок и слухового анализа. Гласные рассматривались в следующих позициях: в ударном, в первом предударном, во втором предударном и в заударном (неконечном слогах). Необходимо отметить, что символами [o] и [e] обозначались как соответствующие ударные гласные, так и появляющиеся в интерферированной речи [o]-образные и [e]-образные безударные гласные.

Особого внимания требует гласный /ɨ/, часто реализуемый дикторами как [ɪ]. Предварительная акустическая картина показала, что характерная особенность нормативного /ɨ/ – изменение частоты FII во времени (начальное значение частоты ниже, чем конечное) – в реализации дикторов-узбеков не прослеживается. Именно поэтому в приведенных ниже результатах анализа гласных начальный участок, реализованный крайне редко, не учитывался.

В результате спектрального анализа были определены значения частоты (в Гц) первой форманты (FI) и значения частоты (в Гц) второй форманты (FII) гласных для спонтанной речи и чтения ФПТ. Затем было проведено сравнение этих параметров.

Известно, что значение FI коррелирует с подъёмом гласного: чем выше её значение, тем ниже подъём гласного, Таким образом, исходя из полученных данных, можно сделать вывод о том, что все исследуемые фонемы, кроме фонемы /e/, в спонтанной речи представлены в более закрытом варианте, чем в чтении.

Вторая форманта коррелирует с рядом, и чем выше вторая форманта, тем ближе гласный к переднему ряду. По значениям FII в чтении и в спонтанной речи можно проследить следующую закономерность: значение FII ниже в спонтанной речи, чем в чтении. Это говорит о том, что в спонтанной речи дикторы реализуют более заднюю артикуляцию рассматриваемых гласных.

На рисунке 1 наглядно представлено расположение гласных на плоскости, где ось значений OX – значения FI, а OY – значения FII.

Отклонение значений формант от нормы в чтении более значительно, чем в спонтанной речи. Вероятно, это обусловлено тем, что при чтении диктор следит за чёткостью дикции, поэтому, и возникает полный тип произнесения, результатом которого служит недостаточная редукция безударных гласных.

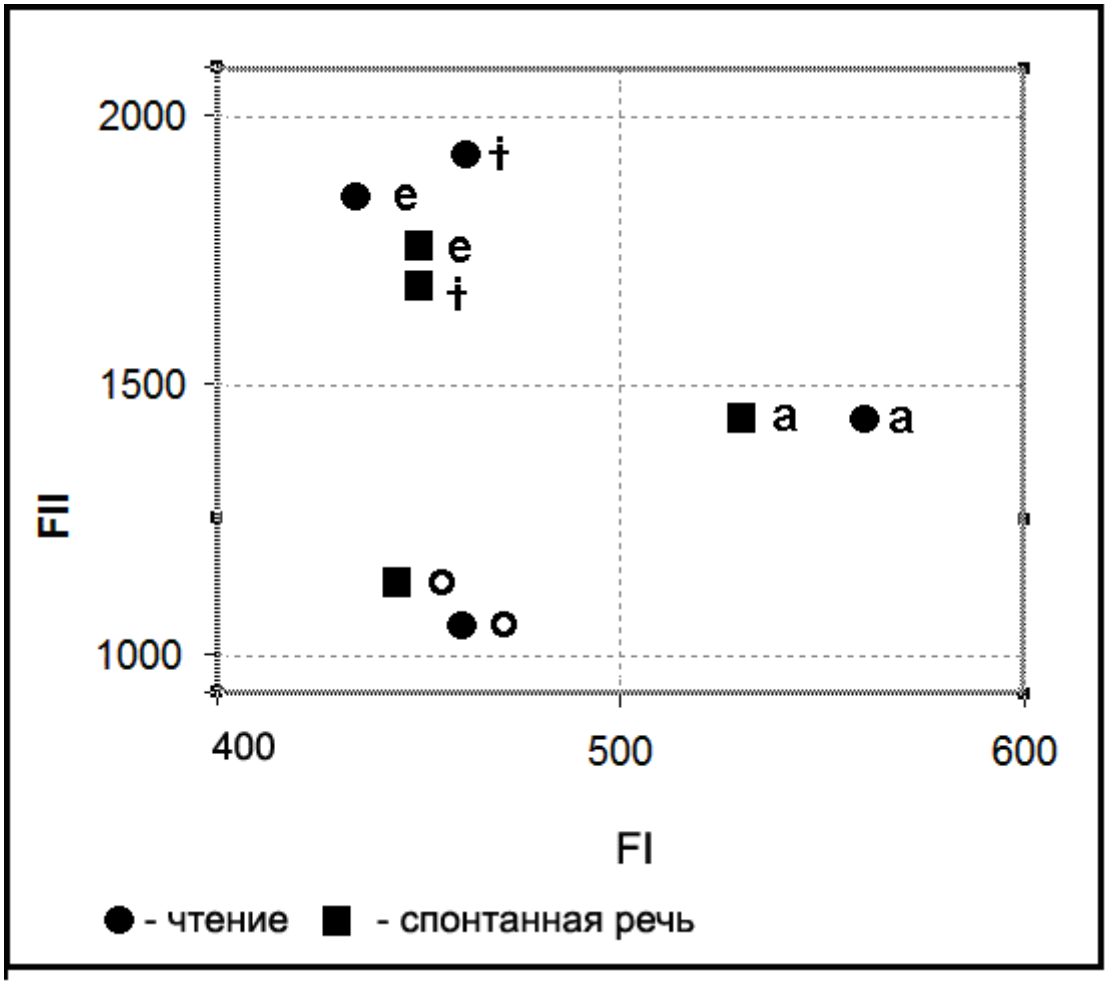


Рис. 1. Сопоставление значений формант ударных гласных, попадающих в область интерференции при чтении ФПТ и в спонтанной речи.

Результаты данного комплексного анализа позволили подтвердить обусловленные взаимовлиянием фонетических систем двух языков отклонения от русской произносительной нормы в речи носителей узбекского языка. Определены следующие особенности:

- особенность речи дикторов, обусловленная их принадлежностью к «окающей» группе говоров, ярко проявляется лишь при чтении дикторами ФПТ, кроме того, в чтении данное явление поддержано орфографическим фактором;
- ввиду недостаточной редукции гласных в безударной позиции существенные для перцепции изменения в речи дикторов-узбеков претерпевают такие гласные русского языка, как /e/, /ɨ/, /a/, /o/. Наиболее ярко данное явление выражено в чтении.

3. Консонантизм

Наряду с «типичными» акцентными ошибками в области консонантизма (недостаточное смягчение мягких, недостаточная веляризованность твердых, отсутствие ассимиляции по глухости-звонкости) в ходе слухового анализа встречались случаи, характерные именно для узбекско-русской интерференции. Так, мягкие согласные приобретали отсутствующую в русской норме фрикативную фазу после размыкания смычки. На месте мягкого /t'/ часто произносится шепелявая аффриката. При апикальной артикуляции согласных в позиции перед гласными переднего ряда можно наблюдать следующую закономерность – чем выше подъём гласного, тем сильнее аффрицируется предшествующий согласный. Непоследовательное появление или отсутствие палатализованных звуков в русской речи узбеков является наиболее частой ошибкой.

Большую трудность для носителей узбекского языка представляет группа зубных согласных. При артикуляции данной группы согласных дикторы-узбеки используют привычную им апикальную артикуляцию. Так, например, русский шипящий /š/, который является какуминальным веляризованным звуком со вторым задним фокусом шума, реализуется без второго заднего фокуса, но со вторым средним. Наиболее частотная замена зубных согласных альвеолярными встречается при реализации согласных /t/, /d/.

Твёрдую аффрикату /c/ в подавляющем большинстве случаев дикторы реализуют как глухой свистящий [s], хотя артикуляция аффрикаты дикторам должна быть знакома по соответствующему сочетанию [t + s] в родном языке. Часто встречающейся ошибкой является так же ярко выраженная невеляризованность этого звука, особенно перед гласными переднего ряда.

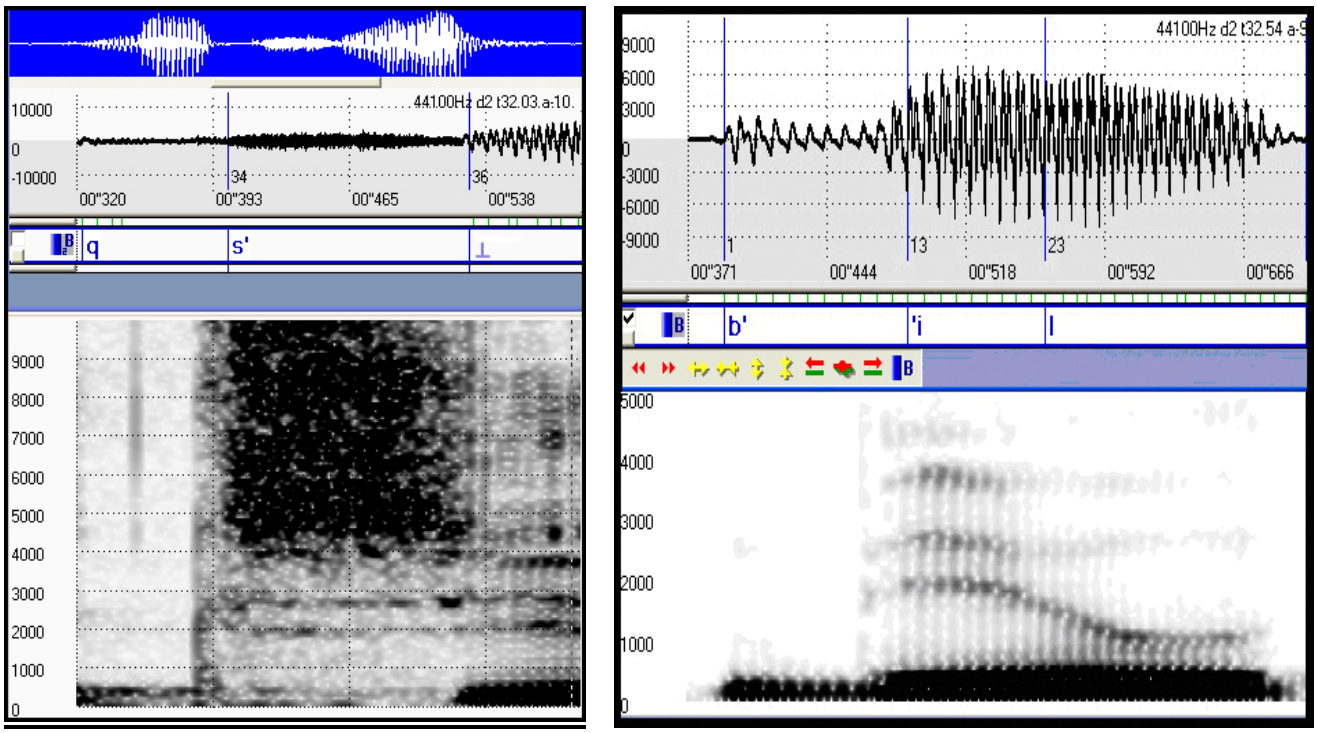
Результаты слухового анализа согласных в спонтанной речи и в чтении ФПТ в целом сходны: согласные в спонтанной речи исследуемых дикторов в сопоставлении с чтением претерпели несущественные модификации, таким образом, различие между значениями ошибочных реализаций согласных в спонтанной речи и чтении статистически несущественно.

В результате проведённого слухового анализа были уточнены зоны интерференции. Для получения ясной картины было необходимо проведение акустического анализа согласных.

Яркими ошибками в русской речи узбеков является замена заднеязычного /k/ на увулярный [q]. На рисунке 2а представлен фрагмент динамической спектрограммы слова «лекциях» в реализации одного из дикторов. На спектрограмме для глухого взрывного [q] можно чётко выделить фазу глухой смычки и высокочастотный короткий шум.

Кроме того, в данном слове интерес вызывает и реализация аффрикаты /c/, которая также входит в область интерференции. Как видно из рисунка, диктор реализовал первую фазу, фазу смычки, данного элемента, возможно, отнеся эту фазу смычки к предшествующему согласному.

Следует отметить и еще один пример реализации [ı] открытого в позиции, где должна быть реализована фонема /ɨ/. Если посмотреть на динамическую спектрограмму слова «был», представленную на рисунке 2б, можно убедиться в непоследовательности палатализации в русской речи узбеков. Как известно, самым существенным показателем палатализации согласного является наличие [i] – образного переходного участка в начале следующего гласного. Такой [i] – образный участок имеется во всех случаях, когда гласный следует за палатализованным согласным, независимо от того, чем характеризуется его палатализация. В слове «был», представленном на рисунке 2б, данный участок ярко выражен, что свидетельствует о наличии мягкого согласного в позиции, где в норме должен был реализоваться твёрдый.



(а) (б)

Рис. 2. Фрагменты осциллограмм и динамических спектрограмм различных слов в реализации одного из дикторов: (а) слово «лекциях»; (б) слово «был».

Нередко в произношении дикторов встречалась и обратная замена: когда на месте мягкого реализуется соответствующий твёрдый согласный. Один из таких случаев мы можем наблюдать на спектрограмме, представленной на рисунке 3.

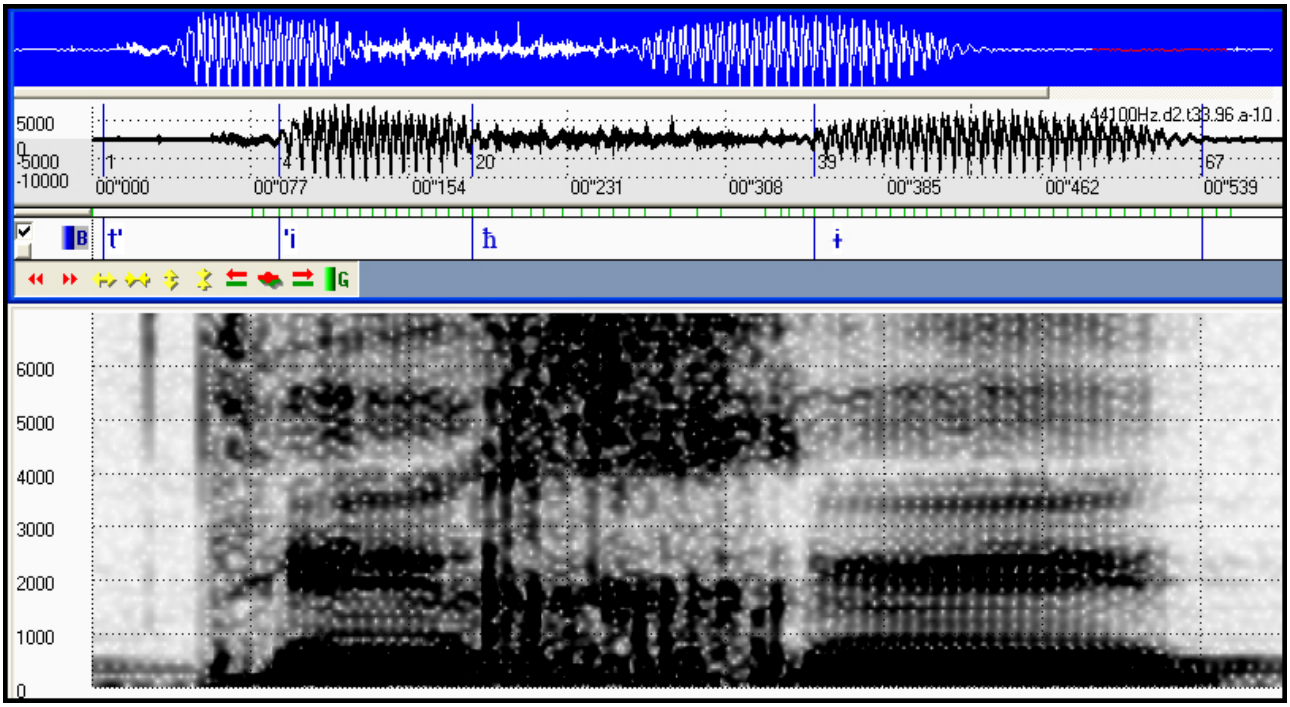


Рис. 3. Фрагмент осциллограммы и динамической спектрограммы слова «тихий» в реализации одного из дикторов.

При слуховом анализе данного фрагмента ярко выраженным отклонением от нормативного произношения является замена заднеязычного /х/ на [ħ] и, как следствие, реализация на месте /i/ гласного более глубокой артикуляции [i̠], реализованного с достаточно ровной FII, что свидетельствует об отсутствии у данного элемента дифтонгоидности, имеющейся при нормативной реализации данного звука.

4. Заключение

Таким образом, можно сделать вывод, что аналогичный описанному выше, но более подробный акустический анализ имеющихся интерферированных звукозаписей позволит выявить акустические признаки сегментного и просодического уровней, обладающие потенциальными диагностическими возможностями, а при наличии достаточного количества записей сделает возможным определение реальных диагностических признаков путем статистической обработки данных.

Необходимо добавить, что метод идентификации по голосу - один из самых удобных для пользователя. Он надежен, так как для диктора сложно качественно имитировать какой-либо акцент, и, кроме того, в исследовании используется мультипараметрический анализ. Предлагаемый метод доступен (возможна удаленная и скрытая идентификация) и дешев (не требует затрат на приобретение дополнительного оборудования), исследуемый материал (звукозапись) можно получить бесконтактным путем.

Литература

1. Бондарко Л. В., Степанова С. Б. Текстовый модуль «Фонетически представительный текст» // Бюллетень Фонетического фонда русского языка, № 14, сентябрь 1992, Bochum; Л., 1992.
2. Болотова О.Б. Гласные в спонтанной речи и при чтении спонтанного текста: автореф. канд. дис. СПб., 2006.

А.В. Венцов

Словесное ударение, собственная длительность гласных и ментальный лексикон*

Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия,
av.ventsov@gmail.com

Принято считать, что обнаружение места словесного ударения является важной операцией в процессе восприятия речи человеком, хотя прямых экспериментальных подтверждений этому, насколько я знаю, не существует.

Применительно к русскому языку, основным признаком словесного ударения считается длительность гласного: в ударном слоге она больше, чем в безударном [1]. Однако одного этого еще недостаточно для обнаружения ударности в непрерывном потоке речи. Необходим алгоритм адекватной оценки ударности конкретного гласного. Можно предположить, что таковым может быть правило «ударным является гласный, длительность которого превышает длительность ближайших соседей на некоторую пороговую величину».

Проверка такого алгоритма на прочитанном в лабораторных условиях материале показала, что при 20-процентном пороге длительность ударного гласного «превышает» длительность ближайших «соседей» не более, чем в 70% случаев [2]. И хотя при такой оценке не учитывалось фонетическое качество гласных, в этой же работе было показано, что длительность ударного «и» всего в 42% случаев оказывается максимальной в пределах фонетического слова.

Очевидно, что при реализации рассматриваемого алгоритма необходимо учитывать эффект собственной длительности гласных: гласные разной степени раскрытия характеризуются разной длительностью в речевом сигнале (самый долгий – [a], самый краткий – [i]). Причем различие это обнаруживается не только при инструментальном анализе речевого сигнала, но и в экспериментах по восприятию словесного ударения [3].

Оказалось также, что равновероятное восприятие ударности в паре [a-i] наступает, когда длительность [a] в 1,5-2 раза превышает длительность [i], а паре [a-u] подобное отношение составляет 1,2-1,5. При разности длительностей в 20% вероятность восприятия ударным более долгого гласного не превышает 25%.

Таким образом, может оказаться, что использование различий в длительностях последовательных гласных в фонетическом слове не приведет к надежной локализации места словесного ударения.

В этих условиях очень важно оценить, какого рода и в каком количестве могут возникнуть ошибки при восприятии звучащей речи, если при определении места словесного ударения опираться только на абсолютные длительности гласных, не учитывая влияние их собственных длительностей, и не позволит ли обращение к ментальному лексикону уменьшить возникающие при этом потери.

Для этого из словаря омографов [4] были выбраны все акцентные омографы типа ***за+мки-замки+, па+рите-пари+те*** и т.п., в которых эффект собственной длительности гласных может привести к наибольшему числу ошибочных решений о месте словесного ударения. Всего было обнаружено 307 таких пар. Затем по полному Национальному корпусу русского языка (www.ruscorpora.ru) были определены частоты обоих элементов каждой пары.

Оказалось, что в 174 парах ни одна из составляющих их словоформ не встретилась в текстах Корпуса, а в 97 – только одна из них была представлена в Корпусе хотя бы один раз. Оставшиеся 36 пар образуют следующие характерные группы:

* Работа выполнена при поддержке гранта РФФИ № 09-06-00244-а.

- в двух парах (да+ли/дали+, ча+сти/части+) наблюдается огромная разница в частоте встречаемости компонентов, 490/2 и 631/1 соответственно;
- восемь пар образованы словоформами, представляющими одинаковые части речи, но в разных временных и падежных формах, и к тому же имеющими разное значение (за+мки/замки+, па+йки/пайки+, ска+чки/скачки+); соотношение частот в этих парах не превышает десяти, а абсолютное значение – 27;
- в тринадцати парах оказались разные части речи (Ва+ри/вари+, запра+вила/заправи+ла, па+ди/пади+);
- еще тринадцать пар образованы глаголами (запла+тите/заплати+те, па+рить/пари+ть, та+щите/тащи+те); частота встречаемости словоформ в этих парах не превышает 15.

Учитывая в целом невысокую частоту встречаемости образующих упомянутые пары словоформ (при объеме Корпуса, превышающем 160 млн. словоупотреблений), можно предположить, что ошибки в интерпретации входного акустического сигнала, вызванные неадекватной локализацией места словесного ударения, могут быть исправлены путем обращения к ментальному лексикону и использования содержащейся в нем грамматической и семантической информации. К тому же, число подобного рода ошибок может оказаться незначительным, а фразовый контекст также поможет снять неопределенность, вызванную ошибочной локализацией словесного ударения.

В заключение следует заметить, что до настоящего времени неизвестны правила и критерии, в соответствии с которыми определяются границы гласных в слуховой системе человека. Соответственно отсутствуют и формальные правила их определения при инструментальном анализе речевого сигнала. В исследованиях, посвященных анализу временных параметров речевого сигнала, обычно не сообщаются правила определения соответствующих границ. Все это делает весьма затруднительным сопоставление количественных результатов, полученных разными авторами.

Литература

1. Физиология речи. Восприятие речи человеком / Отв. ред. Л.А. Чистович.- Л.: Наука, 1976.- 388 с.
2. Кривнова О.Ф. Длительность как средство реализации словесного ударения в тексте (сопоставительный анализ разных способов оценки выраженности ударения в слове) // Язык и речь: проблемы и решения / Под ред. Г.Е. Кедровой и В.В. Потапова.- М.: МАКС Пресс, 2004.- С. 77-99.
3. Kuznetsov V.B., Ott A., Ventsov A.V. Inherent Vowel Duration in Russian: Production and Perception Data // Proceedings of the XIth ICPhS, August 1-7, 1987, Tallinn.- 1987.- Vol. 5.- P. 366-369.
4. Венцов А.В., Грудева Е.В., Касевич В.Б. и др. Словарь омографов русского языка / Ред. Венцов А.В., Касевич В.Б.– СПб.: Филологич. ф-т СПбГУ, 2004.– 160 с.

А.В. Венцов, Е.И. Риехакайнен

Восприятие омофонов и структура ментального лексикона*

Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия,
av.ventsov@gmail.com, reha20@rambler.ru

1. Введение

Ранее было показано [1, 2], что основным фактором, обеспечивающим надёжное распознавание редуцированных словоформ, является контекст. Однако в ряде случаев обращение к нему является невозможным или не позволяет выбрать единственную интерпретацию неоднозначного сигнала. В условиях естественной коммуникации это происходит тогда, когда неоднозначной оказывается самая первая словоформа в высказывании (и опора на контекст невозможна в принципе) и когда левый контекст не позволяет снять неоднозначность. В таких случаях, по-видимому, все единицы перцептивного словаря, активированные входным речевым сигналом, должны сохраняться в памяти слушающего до тех пор, пока контекст не позволит выбрать одну из них. Однако даже при отсутствии достаточного контекста одни словоформы являются более вероятными кандидатами на распознавание, чем другие, т.е. получают более высокую степень активации. Результаты распознавания редуцированных словоформ позволяют считать частотность тем фактором, который влияет на степень активации единиц при отсутствии достаточного контекста [3]. Тем не менее, выдвинутое предположение требует дальнейшей проверки, поскольку словоформы, которые выбирались в качестве ответов в экспериментах по распознаванию редуцированных словоформ, различались не только частотностью, но и степенью соответствия входному речевому сигналу.

В статье будут рассмотрены некоторые результаты исследования, направленного на проверку следующей нулевой гипотезы: при наличии нескольких единиц в перцептивном словаре, в равной степени соответствующих входному сигналу и допускаемых существующим контекстом, преимущество в процессе конкуренции получает та из них, которая является наиболее частотной.

2. Результаты проведенного эксперимента

В качестве экспериментального материала были использованы омофоны — «разные слова, совпавшие по звучанию, при различии морфонологического состава» [4: 288]. В словаре слушающего омофоничным отрезкам должны соответствовать разные по лексическому и/или грамматическому значению единицы, имеющие одинаковую фонемную структуру (т.е. в одинаковой степени соответствующие входному речевому сигналу). Предполагается, что в перцептивном словаре омофоны представлены одной единицей, обнаружение которой в акустическом сигнале приводит к немедленной активации информации следующих уровней (грамматических характеристик, частоты встречаемости) об обоих омофонах [5: 65]. При этом, на наш взгляд, в условиях отсутствия контекста, снимающего неоднозначность, единица, имеющая большую частотность, может быть активирована сильнее.

В качестве экспериментального материала были использованы 4 типа омофонов, одним из которых были возвратные глаголы прошедшего времени единственного числа с безударным окончанием /а/: [sl*ma+l*s’][1] (сломалось-сломалась), [stuča+l*s’] (стучалось-стучалась) и т.п. В рамках статьи будут рассмотрены только результаты, полученные на данном типе омофонов.

* Работа выполнена при частичной поддержке гранта РФФИ № 09-06-00244-а.

[1] Символ «+» в транскрипции указывает на ударность предшествующего гласного. Безударные аллофоны фонемы /а/ обозначаются в фонетической транскрипции знаком «*».

Отбор стимулов для тестовой последовательности осуществлялся с учётом относительной частотности омофонов в паре. Частотность словоформ определялась по полным данным основного корпуса Национального корпуса русского языка [6] (далее — НКРЯ; 162 132 407 словоупотреблений). Безусловно, соотношение текстов в данном корпусе (прежде всего то, что значительная его часть — это художественные тексты, а доля устной речи составляет лишь около 4%), время их создания (с середины XVIII века), а также наличие значительного количества ошибок не позволяют считать, что НКРЯ отражает реальную частоту встречаемости словоформ в русской речи. Однако поскольку некоторые из омофоничных словоформ являются очень редкими, определение их частотности по какому-либо из подкорпусов НКРЯ или по другим корпусам (например, по более сбалансированному Корпусу русского литературного языка, далее — КРЛЯ [7]) не представлялось возможным.

В условиях устно-письменного эксперимента на ответы испытуемых могла повлиять степень соответствия между звучанием словоформы и её написанием, поэтому для каждого из выбранных типов омофонов были подобраны следующие пары:

- омофон, орфографическая запись которого ближе к произношению, является более частотным, чем второй омофон данной пары;
- омофон, орфографическая запись которого ближе к произношению, является менее частотным, чем второй омофон данной пары.

Всего было отобрано 40 пар омофонов, 6 из них относились к типу, рассматриваемому в статье.

К сожалению, необходимый для проведения эксперимента материал невозможно было получить из записей спонтанной речи в силу отсутствия представительных корпусов естественной звучащей русской речи большого объёма. Поэтому тестовая последовательность была составлена из изолированных омофонов, которые были извлечены из предложений, прочитанных двумя дикторами (мужчиной и женщиной). Стимулы предъявлялись испытуемым для прослушивания через наушники. Каждый омофон звучал только один раз.

Испытуемые были разделены на две группы, первой из которых предлагалось просто записать услышанное. При такой инструкции межстимульный интервал составлял 5 секунд, эксперимент продолжался 4 минуты 25 секунд. Подобное задание достаточно близко к диктанту и, следовательно, при его выполнении роль степени соответствия между написанием и произнесением может оказаться значительной.

Задание, предложенное второй группе испытуемых, должно было уменьшить влияние данного фактора: участникам эксперимента предлагалось составить любую фразу с каждым стимулом. Если же они не могли этого сделать, им предлагалось записать хотя бы услышанное слово. Предполагалось, что, выполняя подобную инструкцию, испытуемые будут не просто механически записывать услышанное, а обращать внимание прежде всего на план содержания предъявляемых словоформ.

В данном случае пауза между стимулами составила 20 секунд, а продолжительность эксперимента — 14 минут 25 секунд.

В эксперименте приняли участие 85 человек в возрасте от 15 до 54 лет: 50 в первой группе и 35 — во второй. Испытуемыми были 43 женщины и 42 мужчины (соотношение для первой части 26/24, для второй — 17/18 соответственно).

Все испытуемые были наивными носителями языка: ни один из них не имел лингвистического/филологического образования и не являлся студентом филологического факультета.

Стимулы, которые представляли собой оканчивающиеся на безударный /а/ формы прошедшего времени единственного числа возвратных глаголов, были восприняты испытуемыми очень хорошо: в бланках ответов не встретилось ни одного отказа. При подсчёте результатов был вычислен процент форм женского и среднего рода в ответах испытуемых, а также доверительные интервалы на 5%-ном уровне значимости для всех полученных процентов.

Предположим, что в ментальном лексиконе носителя языка один из разделов образован возвратными глаголами прошедшего времени, а внутри последнего существуют подразделы, объединяющие словоформы, которые совпадают по числу и роду. Тогда можно было ожидать, что оценки этой группы стимулов испытуемыми 1-ой группы примерно поровну будут распределяться между словоформами женского и среднего рода, поскольку в корпусах разного объема количество соответствующих словоформ оказывается примерно одинаковым (см. Таблицу 1).

Таблица 1. Частотность глагольных форм на **-лась/-лось** в корпусах разного объема

	КРЛЯ		**НКРЯ**	
	0,2 млн	*1,05 млн*	*6 млн*	*162 млн*
Частота глаголов на **-лась**	459	2907	19943	447141
Частота глаголов на **-лось**	524	2908	19447	454092
Отношение частот	0,8759	0,9996	1,0255	0,9847

Однако, как следует из Таблицы 2, в 1-ой группе испытуемых по мере увеличения в НКРЯ отношения частоты словоформ женского рода к таковой для среднего рода процент форм женского рода в ответах систематически увеличивается, а процент форм среднего рода — уменьшается. Можно предположить, что в условиях данного эксперимента при оценке данного класса омофонов испытуемые ориентируются на частоту, с которой представлены в ментальном лексиконе соответствующие словоформы.

Таблица 2. Распознавание омофонов на **-лась/-лось**.

Стимул	**Соотношение частот форм ж.р. и ср.р.**		**1-я группа, % ответов**		**2-я группа, % ответов**	
	абсолютн.	*относит.*	*ж.р.*	*ср.р.*	*ж.р.*	*ср.р.*
[sluči+l*s']	1310 / 16743	0,078	6	94	5,7	94,3
[p*luči+l*s']	1544 / 6784	0,227	2	98	5,7	94,3
[pr*d*lža+l*s']	1793 / 2252	0,796	28	72	51,4*	48,6*
[sl*ma+l*s']	263 / 151	1,742	36*	64*	85,7	14,3
[z*krы+l*s']	618 / 99	6,242	44*	56*	77,1	22,9
[stuča+l*s']	63 / 6	10,5	60*	40*	71,4	28,6

*Звёздочкой отмечены те пары значений, которые достоверно не различаются.

Эти данные говорят также о том, что информация о частоте сопутствует представлению каждой словоформы в словаре и что носители языка при необходимости умеют ею пользоваться.

Наличие влияния суммарной частотности классов на **-лась** и на **-лось** можно предположить только для стимулов [sl*ma+l*s'], [z*krы+l*s'] и [stuča+l*s'], для которых количество форм женского и среднего рода в ответах 1-ой группы испытуемых достоверно не различается. Бо́льшая значимость данного фактора по сравнению с частотностью отдельных словоформ именно для данных стимулов может объясняться сравнительно низкой абсолютной частотностью омофонов, входящих в эти пары (см., например, [8]).

В ответах же 2-ой группы испытуемых не наблюдается столь строгой зависимости результатов распознавания от частотности словоформ. Видимо, при необходимости сконструировать фразу с услышанным омофоном испытуемый может опираться не только на частоту встречаемости соответствующей глагольной словоформы, но и на другие факторы.

Так, можно предположить, что в данном случае выбор формы глагола во многом определяется тем существительным или местоимением, которое испытуемый использует в качестве подлежащего при глаголе-стимуле. Например, большее количество форм женского рода в ответах 2-ой группы испытуемых по сравнению с 1-ой может объясняться большей частот-

ностью существительных женского рода по сравнению с существительными среднего рода (как в целом, так и только в именительном падеже): в подкорпусе НКРЯ со снятой омонимией количество существительных женского рода превосходит количество существительных среднего рода примерно в три раза.

Кроме того, значимым фактором может оказаться количество различных существительных и местоимений женского и среднего рода, которые могут употребляться в функции подлежащего при той или иной глагольной форме, а также соотношение их суммарных частотностей. В некоторых случаях, по-видимому, можно говорить о влиянии на выбор формы глагола частоты встречаемости конкретных существительных с конкретными глаголами в пределах одной фразы (например, *дверь закрылась*, *машина сломалась*, *это случилось*, *всё получилось*). Однако все сделанные предположения носят предварительный характер и требуют более детального изучения.

3. Заключение

Проведённый эксперимент в очередной раз свидетельствует о том, что ментальный лексикон слушающего представляет собой сложную многомерную структуру, элементы которой соединены связями, ведущую роль в формировании которых играет частотность.

При этом, несмотря на то, что полученные результаты позволяют сделать ряд интересных выводов и наблюдений, необходимо отдавать себе отчёт в том, что представления о частотности единиц в речи складываются на основе субъективного опыта каждого человека. Следовательно, соотношения частотностей единиц в ментальных лексиконах различных носителей языка могут различаться, и данные, извлечённые из любого, даже представительного и хорошо сбалансированного корпуса, позволяют получить лишь приблизительные представления о той частотности, которой пользуется конкретный носитель языка в процессе распознавания речи.

Литература

1. Риехакайнен Е.И. Стратегии восприятия редуцированных словоформ (на материале русской спонтанной речи) // Научные чтения — 2006. Материалы конференции. 13-14 ноября 2006. С.-Петербург. СПб, 2008. С. 52-59.
2. Венцов А.В., Риехакайнен Е.И., Слепокурова Н.А. Ментальный лексикон и восприятие редуцированных словоформ // Труды первого междисциплинарного семинара «Анализ разговорной русской речи» (АР^3 – 2007). СПб, 2007. С. 21-24.
3. Риехакайнен Е.И. Роль фактора частотности в процессах порождения и восприятия редуцированных словоформ // VII выездная школа-семинар «Проблемы порождения и восприятия речи»: Материалы. Череповец, 2008. С. 53-60.
4. Ахманова О.С. Словарь лингвистических терминов. М., 2007.
5. Венцов А.В. Восприятие устной речи и ментальный лексикон // Русская языковая личность: Материалы шестой выездной школы-семинара. Череповец, 2007. С. 63-69.
6. Национальный корпус русского языка. URL: http://www.ruscorpora.ru
7. Корпус русского литературного языка. URL: http://www.narusco.ru
8. Alegre, M., Gordon, P. Frequency Effects and the Representational Status of Regular Inflections // Journal of Memory and Language 40 (1999). P. 41-61.

Н.А.Слепокурова, Е.П.Комовкина

Структура межпаузальных интервалов спонтанного диалога как фактор взаимодействия собеседников[1]

Санкт-Петербургский государственный университет,
Санкт-Петербург, Россия
nataliars@inbox.ru, ekomovkina@yahoo.com

В 70-80-ых гг. прошлого века в США в рамках социолингвистических исследований появилось направление, получившее название «конверсационный анализ» (conversation analysis), одной из задач которого стало исследование организации устного диалога и, в частности, изучение признаков, отвечающих за так называемый turn-taking – соблюдение очередности реплик участников разговора [1, 2]. Сама по себе данная проблема, сколь бы незначительной она ни показалась на первый взгляд, отнюдь не является тривиальной: как и любое интерактивное взаимодействие, процесс разговора с участием двух или более собеседников должен безусловно подчиняться некоторым системным конвенциям. Не вызывает сомнений, что наряду с существованием неписаных социально закрепленных правил ведения разговора имеется и некоторый набор чисто коммуникативных «ключей», опирающихся, с одной стороны, на некие признаки самого речевого сигнала, с другой – на мимику и жестикуляцию говорящих, ориентируясь на которые участники диалога определяют моменты завершения «порций речи» собеседников и инициирования собственных реплик. Достаточно вспомнить, например, что журналисты, ведущие в теле- и радиоэфире диалоги с удаленными от них коллегами и потому лишенные возможности опираться на невербальные жесты последних, часто вынуждены специально маркировать окончания своих реплик нарочитым произнесением имени собеседника. Реальная важность бессознательного учета вышеупомянутых «ключей» в спонтанном диалогическом общении может быть продемонстрирована не столько его «успехами», сколько как раз «неудачами», демонстрируемыми, например, эпизодическими появлениями в диалогах наложений реплик собеседников.

Достаточно большая литература на Западе посвящена исследованию ответственных за turn-taking чисто лингвистических признаков. Заметную роль здесь сыграло то обстоятельство, что становление конверсационного анализа совпало по времени с появлением таких важнейших на сегодняшний день междисциплинарных языковедческих направлений, как корпусная и дискурсивная лингвистика с их новым и глубоким исследовательским потенциалом, тесно связанным, с одной стороны, с развитием когнитивных исследований, а с другой – с прогрессом в области компьютерных технологий. Базовыми в изучении поверхностной организации устного диалогического дискурса на долгое время стали так называемые дискурсные маркеры – универсальные с типологической точки зрения вербальные и невербальные единицы (в основном, это так называемые дискурсивные слова – особые, чрезвычайно частотные в любом устном дискурсе и слабо изученные в рамках традиционной лингвистики лексемы типа английских *oh, well, so, I mean* или русских *ну, вот, вообще, как бы,* паузы всех типов, а также характерные для устной речи сбои, повторы, наложения слов и др.), которые, по мнению специалистов, в актах устной коммуникации сигнализируют о неких критических, переломных точках дискурса, иначе говоря, используются для его квантования и структурирования [3, 4].

Именно с этой точки зрения и трактовалась в основном обозначенная выше проблема: появились работы, в которых делались попытки в качестве сигналов к переходу «слова» в диалоге к следующему диктору рассматривать разные типы дискурсных маркеров, однако в целом поиск положительных показателей такого рода, маркирующих момент, когда в разговор может вступить следующий собеседник, оказались не слишком внятными – за

[1] Работа выполнена при поддержке гранта РФФИ № 9-06-00244а.

исключением, разумеется, тех достаточно очевидных случаев, например вопросно-ответных единств, когда сигналом к проговариванию ответа является семантико-просодическая завершенность вопроса [3].

Накопленный нами опыт изучения спонтанных диалогических текстов заставляет считать, что наряду с «положительными» маркерами, разрешающими следующему говорящему вступать в диалог, в речи говорящих, возможно, содержатся и некоторые «отрицательные» маркеры, предупреждающие о том, что данная реплика не закончена и собеседнику следует воздержаться от немедленного начала собственной «порции» речи. Мы попытаемся показать это на примере анализа аудиозаписи радиоинтервью, транслированного по каналу «Радио России, Санкт-Петербург» под рубрикой «Утренний гость». Данный цикл радиопередач проходит по одной и той же схеме: корреспондент интервьюирует очередного гостя и, если вопросы первого, по крайней мере – частично, готовятся заранее и реализуются в весьма сжатой форме, то ответы второго являются, как правило, развернутыми спонтанными репликами.

Указанная аудиозапись, длительность которой составила около 25 минут, с помощью специальной компьютерной программы была подвергнута процедуре сплошной линейной сегментации, в процессе которой исследователь, прослушивая через головные телефоны участок речевого сигнала и одновременно просматривая на мониторе его осциллограмму (а в случае необходимости – и синхронизированную с осциллограммой динамическую спектрограмму), отмечал курсором моменты начала и завершения всех участков сплошного, беспаузального «говорения» и таким образом фиксировал в данном тексте абсолютно все паузы за исключением интервала смычки глухих смычных согласных, а также заносил в компьютер подробный орфографический транскрипт прослушанного отрезка речи; начала и концы межпаузальных интервалов и самих пауз фиксировались при этом компьютером в процессе сплошного хронометрирования[2]. В результате весь исследуемый устный текст был расчленен на последовательные участки «говорения», перемежающиеся паузами, самыми регулярными из которых были паузы вдоха (в размещенных ниже примерах они обозначены символом inh). Отмечались и другие виды пауз – прежде всего заполненные и незаполненные паузы хезитации, вздохи, чмоки, глоттализации и т.п. Указанные типы пауз могли, разумеется, пересекаться: см. замечание Т.М.Николаевой о том, что «более компетентные говорящие стремятся «привести паузы хезитации к дыхательным паузам» (цит. по [5: 68]). Однако на данном этапе изучения межпаузальных интервалов в диалоге различия между типами пауз во внимание не принимались.

Помещенный ниже пример в таблице 1 представляет собой небольшой фрагмент сплошной компьютерной сегментации радиотекста. В левом его столбце последовательно представлены орфографические транскрипты отрезков речи, заполняющих межпаузальные интервалы, а также сами паузы; в крайнем правом столбце зафиксированы длительности пауз и межпаузальных интервалов, два столбца цифр в середине обозначают соответственно моменты их начала и конца на линейной оси времени (все цифры указывают соответственно на минуты, секунды и миллисекунды, прошедшие от начала записи).

Каждый из межпаузальных участков (их оказалось около 450), содержавших произнесенные интервьюируемым диктором-мужчиной отрезки языковых единиц, затем оценивался с точки зрения того, можно ли его считать полноценной синтагмой, т.е. просодически цельнооформленной семантико-синтаксической единицей. В результате анализа было обнаружено, что около трети от числа всех межпаузальных интервалов исследованного текста, представляют собой синтагматически «ущербные» отрезки речи, не подпадающие под определение щербовской синтагмы [6]. Этот факт вступает в противоречие со сложившимися в фонетической практике традиционными представлениями, согласно которым основной единицей устной речи, характеризующейся отсутствием внутренних пауз, является именно синтагма [7]. Эта коллизия, по-видимому, имеет только одно объяснение:

[2] Пользуясь случаем, выражаем глубокую благодарность нашей коллеге Е.И. Корешковой, проделавшей эту разметку устного текста.

понятие синтагмы выводилось на материале устной, так называемой «лабораторной» речи, образцами которой чаще всего служили прочитанные вслух письменные тексты, но отнюдь не спонтанная речь.

Таблица 1. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
никто	01:09,372	01:09,836	00:00,464
inh	01:09,836	01:10,167	00:00,332
не думал создавать коммуну и-и делать какую-то вот такую организацию. Так	01:10,167	01:14,969	00:04,802
inh	01:14,969	01:15,633	00:00,664
случилось	01:15,633	01:16,627	00:00,994
э-э-э	01:16,627	01:17,120	00:00,492
по обстоятельствам я работал	01:17,120	01:18,855	00:01,736
pause	01:18,855	01:18,995	00:00,139
в многострадальном пятьдесят пятом	01:18,995	01:20,759	00:01,764
inh	01:20,759	01:21,361	00:00,602
специнтернате	01:21,361	01:23,264	00:01,902
как его тогда называли	01:23,264	01:24,452	00:01,188
потом он стал	01:24,452	01:25,105	00:00,654
inh	01:25,105	01:25,582	00:00,477
двадцать пятой коррекционной	01:25,582	01:26,961	00:01,379
pause	01:26,961	01:27,380	00:00,418
э-э-х	01:27,380	01:27,896	00:00,516
школой	01:27,896	01:28,452	00:00,556
pause	01:28,452	01:28,639	00:00,187
inh	01:28,639	01:29,696	00:01,056

Вопрос о том, почему реплики собеседников в спонтанных диалогах могут разбиваться паузами в «неправильных» с семантико-синтаксической точки зрения местах – что, кстати, совершенно не мешает самим участникам спонтанной коммуникации, совершенно не замечая этих «неправильностей», прекрасно понимать друг друга, – требует, разумеется, дальнейшего тщательного изучения. В качестве одного из возможных ответов на него мы предположили, что «неправильное» завершение межпаузального интервала в спонтанной диалогической речи может играть роль своего рода отрицательного дискурсного маркера, оформляющего явную незаконченность текущего сообщения и потому предостерегающего собеседника от возможности принять следующую паузу за сигнал к началу развертывания собственной реплики.

Ниже представлена количественно и проиллюстрирована примерами бóльшая часть типов нарушений синтаксических связей, способных, на наш взгляд, выполнять в диалоге функции своеобразных дискурсных маркеров. Напоминаем, что эти нарушения были выявлены в итоге сплошного анализа межпаузальных интервалов в исследованной нами аудиозаписи.

1) В 14% всех рассмотренных межпаузальных отрезков наблюдаются случаи отрыва от глагола-сказуемого дополнений и обстоятельств, заполняющих его обязательные валентности, либо отрыва от глагола-связки присвязочной именной части (таблицы 2, 3).

Таблица 2. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
это значит	07:36,198	07:36,844	00:00,646
что он завтра будет нести в себе	07:36,844	07:39,084	00:02,240
inh	07:39,084	07:39,638	00:00,554
вот эти проблемы	07:39,638	07:40,578	00:00,940

Таблица 3. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
вот я не мог оставить тех детей	02:30,288	02:31,984	00:01,696
которые	02:31,984	02:32,474	00:00,490
inh	02:32,474	02:33,369	00:00,895
ну	02:33,369	02:33,601	00:00,232
э-э-э	02:33,601	02:33,881	00:00,280
сохранились/ так скажем/ раз оказались	02:33,881	02:36,628	00:02,746
э-а-а	02:36,628	02:37,375	00:00,748
относительно сохранные / которым	02:37,375	02:38,991	00:01,616
э-э-э	02:38,991	02:39,464	00:00,472
можно и нужно было помочь	02:39,464	02:40,906	00:01,442

2) 13 % рассмотренных отрезков демонстрируют случаи, когда подлежащие или их группы отделяются паузами от сказуемых (примеры приведены в таблицах 1, 4, 5).

Таблица 4. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
пропущенные блоки социального развития/ воспитания	07:28,203	07:30,910	00:02,706
inh	07:30,910	07:31,520	00:00,610
э-э-э	07:31,520	07:31,859	00:00,340
как бы-ы	07:31,859	07:32,496	00:00,636
э-э	07:32,496	07:32,788	00:00,292
у человека прошли как-то неоткорректированными	07:32,788	07:35,883	00:03,095

Таблица 5. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
что мы	17:05,092	17:05,514	00:00,422
inh	17:05,514	17:06,000	00:00,486
сегодня как бы-ы не	17:06,000	17:07,777	00:01,777
э-э-э	17:07,777	17:08,448	00:00,670
отвечаем современным требованиям/ поэтому	17:08,448	17:10,692	00:02,244

3) В 11% случаев причиной появления «ущербных» в синтаксическом отношении отрезков является изолированное произнесение предлога или союза (редко), либо завершение отрезка оторванным от правого контекста предлогом или союзом (в подавляющем большинстве случаев) (примеры приведены в таблицах 6 – 10):

Таблица 6. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
выставленные	01:46,173	01:47,027	00:00,854
(вздох)	01:47,027	01:47,532	00:00,505
э-э-э	01:47,532	01:48,146	00:00,614
за-а	01:48,146	01:48,553	00:00,407
(вздох)	01:48,554	01:48,839	00:00,286
порог и-и	01:48,839	01:49,723	00:00,884

Таблица 7. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
как мы сейчас вот такой термин нашли	03:06,412	03:08,624	00:02,212
когда	03:08,624	03:09,048	00:00,424
inh	03:09,048	03:09,523	00:00,476
решаются все основные проблемы	03:09,523	03:11,705	00:02,182

Таблица 8. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
ну-у-у во-первых многое изменилось в законодательстве	17:17,244	17:18,490	00:01,246
значит по	17:18,490	17:21,368	00:02,878
э-э-э	17:21,368	17:21,900	00:00,532
защите прав детей	17:21,900	17:23,029	00:01,129

Таблица 9. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
ну вселяет как бы (вздох) надежду	04:50,624	04:52,719	00:02,096
э-э-э	04:52,719	04:53,344	00:00,624
в нас на то что	04:53,344	04:54,375	00:01,032
inh	04:54,375	04:54,923	00:00,548
наше дело	04:54,923	04:55,707	00:00,783
inh+чмок	04:55,707	04:56,447	00:00,740
правое и мы	04:56,447	04:57,816	00:01,369

4) В 8% случаев причиной появления «ущербных» отрезков оказался отрыв определения от определяемого слова. Иллюстрации этого явления содержатся в примерах, приведенных в таблицах 1 и 10, 11.

Таблица 10. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
поэтому я понимаю	05:35,950	05:36,962	00:01,012
что это условные	05:36,962	05:38,350	00:01,388
(вздох)	05:38,350	05:38,537	00:00,188
так скажем направления	05:38,537	05:39,728	00:01,190
нельзя реализовать всю	05:30,054	05:31,304	00:01,250
э-э-э	05:31,304	05:33,000	00:01,696
государственную политику	05:33,000	05:34,405	00:01,406
inh	05:34,405	05:34,889	00:00,484
в отношении семьи	05:34,889	05:35,950	00:01,060

Таблица 11. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
дети подростки имеющие серьёзные	08:50,776	08:52,945	00:02,170
inh	08:52,944	08:53,421	00:00,478
проблемы	08:53,421	08:54,015	00:00,594
э-э-э	08:54,014	08:55,436	00:01,422
ну разного типа	08:55,436	08:56,563	00:01,128

4) 7% межпаузальных отрезков составляют случаи изолированного произнесения дискурсивных слов (таблица 12).

Таблица 12. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
э-э друзья товарищи находят	08:30,148	08:32,208	00:02,060
э-э-э	08:32,208	08:32,528	00:00,320
как бы	08:32,528	08:33,012	00:00,484
(вздох)	08:33,012	08:33,178	00:00,166
уже сами коммунары да?	08:33,178	08:34,444	00:01,266

5) В 5% случаев «неправильные» отрезки образуются из-за разрыва паузой частей единой именной группы (таблица 13).

Таблица 13. Пример сплошной компьютерной сегментации радиотекста

Орфографические транскрипты отрезков речи и пауз	Момент начала сегмента, мин:с,мс	Момент окончания сегмента, мин:с,мс	Длительность сегмента, мин:с,мс
сейчас вот при помощи	20:02,226	20:03,496	00:01,269
inh	20:03,496	20:04,081	00:00,585
наших друзей	20:04,081	20:04,928	00:00,847

Приведенные выше данные, по-видимому, достаточно отчетливо и полно демонстрируют количество и разнообразие обнаруживаемых в спонтанно-диалогической речи «ущербных» с точки зрения классического синтаксиса межпаузальных интервалов. Эти выводы в целом поддерживаются и опытом работы с другими находящимися в нашем распоряжении спонтанными диалогами. Вообще, по нашим наблюдениям, подтверждаемым и данными других исследователей [8], нет таких видов синтаксических зависимостей, которые, пусть и редко, не могли бы подвергаться в спонтанной речи ослаблению за счет появления

пауз между взаимосвязанными словоформами. Обилие пауз в синтаксически «нештатных» позициях, казалось бы, должно не только свидетельствовать о незавершенности высказывания говорящего, но и создавать для слушающего определенные трудности в восприятии, связанные с появлением многочисленных моментов «ожидания продолжения», тем не менее на поверхностном уровне спонтанно-диалогической реальности эти трудности не осознаются и не мешают говорящим без труда понимать друг друга. В чем причина этого – в бессознательном ли «невыходе» говорящих за пределы некоторого критического диапазона вариации пауз по длительности, в компенсирующей ли все сегментные «рытвины и ухабы» роли супрасегментных средств, и прежде всего мелодики, или, наконец, в решающей и подавляющей весь фонетико-грамматический «мусор» функции семантики и прагматики – покажут лишь дальнейшие кропотливые и трудоемкие исследования.

Литература

1. Sacks H., Schegloff E.A., Jefferson G. A simplest systematics for the organization of turn-taking for conversation. Language, 50, 1974. pp. 696-735.
2. Schegloff E.A. Sequence Organization in Interaction: A Primer in Conversation Analysis. V. 1, Cambridge: Cambridge University Press, 2007.
3. Clark H.H. Pragmatics of language performance. In: L.R.Horn & G.Ward (Eds.) Handbook of pragmatics. Oxford: Blackwell, 2004. pp. 365-382.
4. Fraser B. What are discourse markers? Journal of Pragmatics, 31, 1999. pp. 931-952.
5. Кибрик А.А., Подлесская В.И. (ред.) Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М., 2009.
6. Комовкина Е.П., Слепокурова. Н.А. Анализ межпаузальных интервалов в спонтанном тексте: предварительные результаты //"Череповецкие научные чтения - 2009: Материалы Всероссийской научно-практической конференции". Череповец: ГОУ ВПО "Череповецкий государственный университет", 2010. С. 47-51.
7. Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики. СПб., 2000.
8. Степанова С.Б. Соотношение синтагматического и хезитационного членения в спонтанной устной речи. Доклад на конф. «Инновации в исследованиях русского языка, литературы и культуры». Пловдив, 2006. Рукопись.

М.О.Пономарь

Обработка речи на фонетическом уровне в интересах скрытного встраивания данных

Московский государственный лингвистический университет
Москва, Россия,
oponomar@inbox.ru

1. Введение

Использование обработки речи на фонетическом уровне позволяет скрытно внедрять информацию в речь путем такой модификации электроакустического речевого сигнал, которая не обнаруживаются в канале связи и на выходе системы ни на слух, ни с помощью инструментальных средств без сравнения с эталоном речи, которым владеет только передающая сторона. В представленных материалах рассматривается практический аспект этого подхода, заключающийся в оценке возможности технической реализации просодических методов сокрытия информации в речи. Это дает возможность оценить подходящие для сокрытия информации параметры просодии и сформулировать требования к практическим алгоритмам для специализированных цифровых интегральных схем и вычислителей общего назначения. Обоснование этих требований позволяет произвести отбор существующих и наметить разработку новых алгоритмов как для целей скрытой связи по открытым каналам, так и для целей создания аутентификационных меток для фонограмм.

Развитие речевой стеганографии на основе вариативности просодии со временем может приблизиться к тому рубежу, который отделит область научных исследований от разработки опытных и промышленных образцов. Уже в настоящее время в интересах скрытой связи разработан метод и опробован в лабораторных условиях алгоритм встраивания информации в некоторые параметры просодии [1]. Однако имеется потребность в увеличении количества используемых параметров и использования их комбинаторных возможностей в интересах увеличения пропускной способности скрытого канала связи. Многообразие параметров просодии и сложность их формализации являются препятствием для дальнейшего развития методов и практической реализации алгоритмов этого направления защиты информации. В связи с этим необходимо обосновать требования к просодическим стегоалгоритмам и их составным частям, провести отбор и оценить возможности их технической реализации.

2. Состав параметров просодической стегосистемы

Основой вариативности просодии является инвариант, позволяющий вносить в него такие изменения, которые не выходят за пределы допустимого отклонения от психоакустической нормы, а потому не заметны постороннему наблюдателю. Это вариант незаметного для нарушителя канала связи ресинтеза речи. Для того, чтобы использовать просодию в качестве стегоконтейнера, потребуется решить три основных задачи [2].

Во-первых, установить, какие из наблюдаемых параметров просодии являются управляемыми. То есть, для каждого параметра оценить, существуют ли в настоящее время технические и/или программные средства, позволяющие его инструментально измерять и модифицировать. Во-вторых, разработать метод внедрения скрываемых данных в управляемые просодические параметры. И, наконец, установить допустимые пределы модификации просодических параметров, при которых соблюдается требование скрытности. Принципиальные решения этих задач пока найдены только применительно к двум параметрам просодии – частоте основного тона и длительности сегментов речи [1,3,4].

В соответствии с основами традиционной фонетики [5,6] материальными акустическими средствами языковой просодии являются параметры, приведенные в

таблице 1. (в таблицу не включены некоторые супрасегментные речевые средства, такие, как артикуляция, эмоциональная окраска речи и другие, не имеющие пока средств для инструментальной оценки).

Таблица 1. Акустические средства языковой просодии

№	Группа	Параметр	Изменчивость	Размерность
P1	слоговые тоны	уровень тона	низкий-высокий	Гц
P2		контурность тона	восходяще-нисходящий	дб/с
P3	ритмические схемы слова и фразы	интенсивность основного ударения	меньше-больше	дб
P4		интенсивность второстепенного ударения	меньше-больше	дб
P5		длительность ударных гласных	меньше-больше	мс
P6		длительность безударных гласных	меньше-больше	мс
P7		длительность пауз	меньше-больше	мс
P8		расположение пауз	меньше-больше	мс
P9	фразовые тоновые интонации	уровень акцента	низкий-высокий	Гц
P10		длительность акцента	меньше-больше	мс
P11		интенсивность акцента	меньше-больше	дб

На примере образца речи (рисунок 1), обработанного в системе PRAAT [7] (в верхней части - волновая форма, в нижней – кривые частоты основного тона и интенсивности), показан подход к акустическому измерению параметров просодии. Инструментальная доступность для измерения является необходимым условием для использования параметра просодии в качестве стегоконтейнера. Таким образом, параметрами-контейнерами могут стать: уровень, контурность тона; интенсивность основного, второстепенного ударения; длительности и расположение пауз; уровень, длительность и интенсивность фразового акцента (P1-P11).

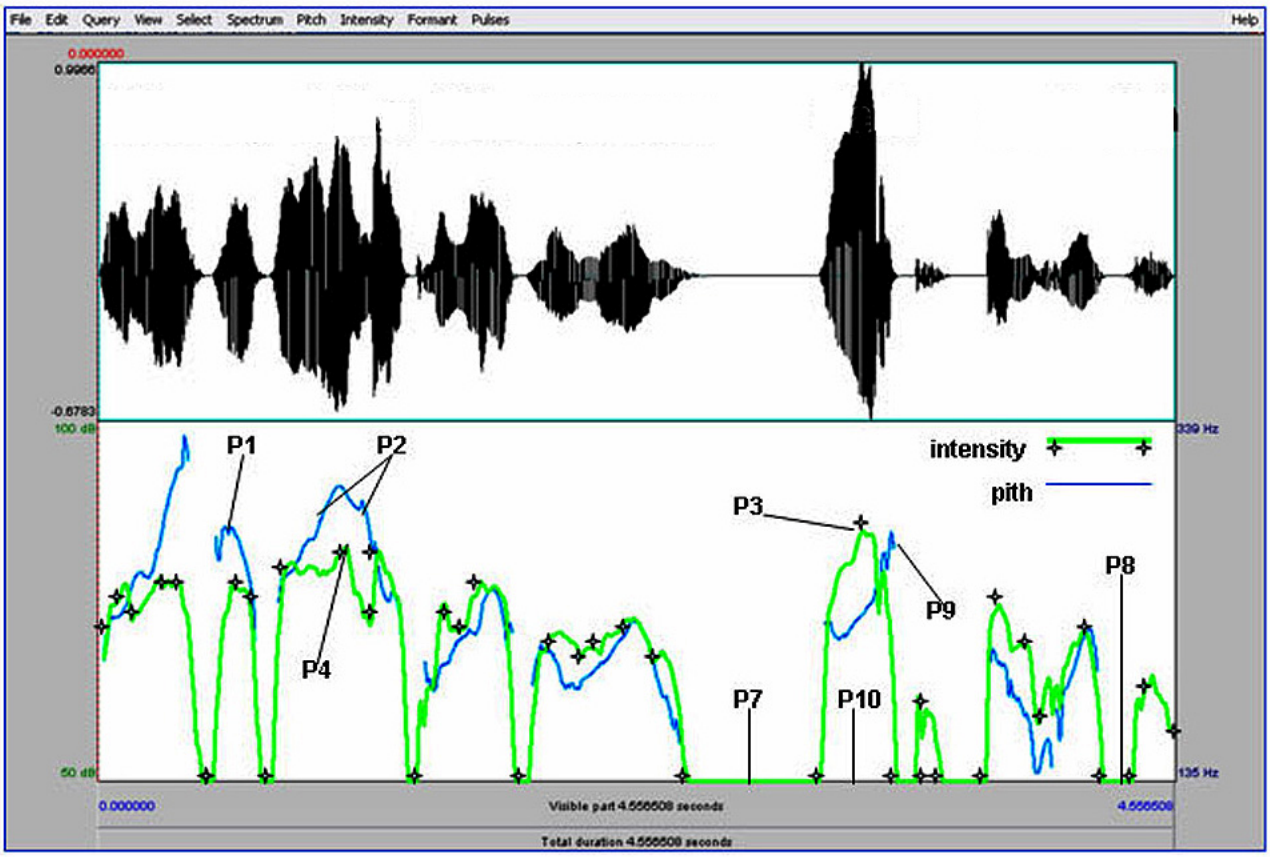


Рис. 1. Возможности измерения параметров просодии.

3. Обоснование требований и оценка эффективности алгоритмов

Все перечисленные выше параметры являются физическим отражением просодических свойств речевого сигнала, не связанных непосредственно с его цифровый формой при передаче, обработке и хранении. Поэтому к ним применим общий подход к встраиванию информации и типовая последовательность алгоритмов: а) сегментации речевого потока; б) определения значения несущего параметра в каждом сегменте; в) вычисления нового его значения при помощи QIM –кодирования и шифрования; г) его модификации [1-4,8]. Однако сложность проектной разработки, технической реализации и область их применения (связь (С) и/или аутентификация (А)) существенно различаются. Например, для длительности паузы (Р7) все преобразования в сторону ее увеличения достаточно просты и могут быть выполнены в реальном времени в сеансе связи. В то же время обработка фразовых интонаций (Р9-Р11) возможна только по завершении фразы. Возникающая при этом задержка неприемлема как с точки зрения скрытности, так и коммуникации. В разной стадии проработки находятся и методы сегментации, измерения, модификации параметров (таблица 2.), различны их проектная и вычислительная сложность.

Таблица 2. Сегментация, измерения, кодирование и модификация параметров просодии

парамеры/ алгоритмы	P1	P2	P3	P4	P7	P8	P9	P10	P11
сегментация	+	+	+	?	+	+	+	+	+
измерение	+	?	+	?	+	+	+	+	+
кодирование	+	+	+	?	+	+	+	+	+
модификация	+	?	+	+	+	?	+	+	+
область исп.	С,А	?	С,А	А	С,А	А	А	А	А

В приведенной таблице знаком «+» обозначено наличие, по крайней мере, формализованного подхода, модели, а иногда и программной реализации (часто без доступа к исходному коду) алгоритма. Знаком «?» - или полное отсутствие такого подхода (в фонетической литературе обозначаемое как «очень сложно»), или возможность его реализации только в условиях полного отказа от управления другими параметрами.

Квантование речевого сигнала по времени и уровню является условием встраивания цифровых данных в просодические параметры. Квантование по времени - это сегментация речи на некоторые однородные временные участки. На них измеряемый параметр (тон, интенсивность, длительность, расположение) может быть описан простой функцией или правилом (звук-тишина, тональный-шумовой, подъем-спад, постоянство), а на границах этих участков происходят резкие изменения акустических характеристик. Очевидно, что для каждого параметра может быть разработано множество алгоритмов сегментации. Основное требование к ним – однозначность разбиения на сегменты до и после внедрения скрываемой информации и прохождения по каналу связи. Причем при одновременном, т.н. векторном внедрении данных в несколько параметров речевой поток или фонограмма может состоять одновременно из несколько наборов сегментов. Проектная и вычислительная сложность этих алгоритмов невелика. В настоящее время некоторые их них реализованы в средах C^{++}, Matlab, аудиоредакторе Melodyne ©Celemony Software GmbH для разбиения речевого потока на синтагмы.

Измерение в метрических единицах каждого параметра в сегменте является основой для его QIM-кодирования по уровню, являющегося основой внедрения данных [3] . Из всех параметров, перечисленных в таблице 1., только два имеют размерность Гц (P1,P9) - частоты основного тона (ЧОТ). Но именно определение ЧОТ является наиболее технически сложным в проектировании и реализации. Как указывается в [6], на то есть объективные причины – периодичность вокальных звуков со строгой физической точки зрения достаточно условна. Тем не менее, число разработок определителей ЧОТ постоянно растет, что связано с потребностями вокодерных технологий, распознавания и синтеза речи. Поэтому для этого параметра есть возможность выбора алгоритма. Главное требование – минимальное время вычислений при приемлемой точности, так как вычислительные затраты изначально достаточно большие. Для остальных параметров проектная и вычислительная сложность алгоритмов измерений невелика.

Кодирование с квантованием просодических параметров (QIM) описывается достаточно простым в реализации алгоритмом, основанном на использовании заранее подготовленных кодовых таблиц [3]. Для обеспечения стеганографической стойкости алгоритм QIM комбинируется с поточным криптографическим алгоритмом [8]. Оба алгоритма могут быть эффективно реализованы в специализированных цифровых интегральных схемах.

Модификация просодических параметров тесно связана с их измерениями и имеет те же проблемы применительно к ЧОТ, а также дополнительные проблемы с модификацией длительности вокализованных сегментов. В настоящее время установлено [9], что с практической точки зрения наиболее целесообразно модифицировать такие просодические характеристики, как частота основного тона, интенсивность и длительность отрезков речевого сигнала, непосредственно модифицируя акустический сигнал как таковой, не используя параметрических моделей.

4. Заключение

Использование просодических параметров речи для скрытого встраивания информации является достаточно сложной задачей не только потому, что человеческий слух довольно точно определяет признаки искусственности речи. Эта проблема решается определением пределов психоакустической нормы модификации параметров, обеспечивающей гарантированную скрытность. Необходимо оценить реализуемость алгоритмов и

стегосистемы в целом. Приведенные данные показывают, что в настоящее время для встраивания информации в основном используются исследовательские алгоритмы, разработанные для других применений, что ограничивает их практические возможности. Недостаточно изучены вопросы искусственного воспроизведения контурных тонов, особенностей артикуляции, ритмических схем ударения, средств артикуляции, интонации и других характеристик просодии в интересах ресинтеза речи. Это объясняется, в том числе и тем, что разработка математических моделей просодии, связывающих характеристики просодии в её классическом понимании с материальными, акустическими показателями звуков речи до настоящего времени не имела такого важного практического применения, как задача защиты информации. И не только защиты информации, но и клонирования речи.

Литература

1. Ponomar, Marina. Data hiding in speech signals on the basis of the modification of segment pitch and duration // 19th International Congress on Acoustics ICA2007MADRID, 2-7 Sept. 2007, Madrid, Spain, 2007, CAS-03-023, p.46-49.
2. Пономарь М.О. Использование вариативности речевой просодии при создании интеллектуальных систем защиты информации // Материалы III Всероссийской конференции студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации», 11-13 ноября 2009 г., Москва, МИРЭА, С. 374-377.
3. Пономарь М.О. Кодирование с квантованием несущих параметров речевых сигналов для скрытого встраивания данных // Сборник трудов XVII Международной конференции «Информатизация и информационная безопасность правоохранительных органов». - М.: Академия управления МВД России, 2008, С. 394-396.
4. Ponomar M.O. On Acceptable Modification Limits of Electroacoustic Speech Signals for Data Hiding // Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2009, September 12-14, 2009, Kyoto, Japan), IEEE Computer Society, Los Alamitos, CA, USA, 2009, pp. 551-554.
5. Потапова Р.К. Речь: коммуникация, информация, кибернетика: Учебное пособие. Изд. 2-е, доп. – М.: Эдиториал УРСС, 2001, С. 276-285.
6. Кодзасов С.В., Кривнова О.Ф. Общая фонетика: Учебник. М.: Рос. гос. гуманит. ун-т. 2001, С. 183-194.
7. Paul Boersma & David Weenink. Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/ .
8. Пономарь М.О. Обеспечение стеганографической стойкости при встраивании данных в несущие параметры речи // Труды Всероссийской конференции "Проведение научных исследований в области обработки, хранения, передачи и защиты информации", Ульяновск, УлГТУ, 1-5 декабря 2009г., секция 4 т. 4, С. 65-68.
9. Бабкин А.В. Особенности применения технологии TD-PSOLA для модификации характеристик вокальных аллофонов // Труды международного семинара Диалог’2000 по компьютерной лингвистике и ее приложениям. Москва, 2000.

С.И. Николенко, М.Л. Кореневский, И.А. Пономарёва, К.Е. Левин

Двухпроходное распознавание на основе классификации речи по тематике

«Центр речевых технологий»,
Санкт-Петербург, Россия,
nikolenko@speechpro.com

1. Введение

Одной из актуальных задач обработки речевой информации является автоматическая классификация звуковых файлов, содержащих речь, по тематике того, что было произнесено. Решение задачи классификации может найти прямые применения - категоризация новостных сообщений, диалогов с операторами колл-центров и т.д. Однако в настоящем докладе пойдёт речь о применении классификатора речевых звукозаписей для улучшения качества распознавания речи.

К сожалению, существующие алгоритмы распознавания речи с большим словарём, особенно имеющиеся в настоящий момент разработки для русского языка, имеют ряд серьёзных недостатков. При относительно невысоком быстродействии имеющиеся декодеры допускают большое количество ошибок; особенно высок процент ошибок в системах, работающих в телефонном канале при низком отношении сигнал/шум. Поэтому в промышленных применениях распознавания и классификации речи, например, анализе работы колл-центров, размер словаря приходится значительно ограничивать. Таким образом, классическая задача разделения текстовых сообщений по темам усложняется новыми задачами: как выбрать оптимальный словарь для распознавания и как проинтерпретировать полученные результаты невысокого качества.

В предыдущих работах мы совместили классические методы категоризации текстов с системой автоматического распознавания речи и оценили результаты работы классификаторов на достаточно представительной базе звукозаписей. Эксперименты показали, что даже в условиях низкого качества распознавания речи (уровень ошибок первого и второго рода более 40%) качество классификации полученных результатов оказывается удовлетворительным; результаты этих экспериментов кратко описаны в разделе 2, а также были представлены в [1]. В настоящей работе сделан следующий шаг: мы представляем результаты двухпроходного распознавания речи. Раздел 3 посвящён описанию схемы двухпроходного распознавания и нашим результатам в этом направлении. Выясняется, что качество распознавания речи, особенно количество ложных срабатываний, можно значительно улучшить за счёт классификации промежуточных результатов и уточнения словаря распознавания.

2. Классификация текстов и речи: подходы и результаты

Методы классификации (категоризации) текстов хорошо изучены и имеют богатую историю, начинающуюся с работ 1960-х гг. по наивным вероятностным моделям классификации текстов. В настоящее время промышленно применяются наивные байесовские классификаторы, геометрические классификаторы, такие, как метод ближайших соседей и классификатор Роккио, метод опорных векторов (SVM) и другие вероятностные модели [2]. Все они дают устойчиво хорошие результаты в различных ситуациях применения. Нами применялся очевидный способ классификации речевых сообщений – получение полной текстовки средствами автоматического распознавания речи, а затем применение к полученному тексту алгоритмов классификации. В этом разделе мы кратко описываем несколько алгоритмов классификации, а затем приводим результаты сравнения.

1. *Наивный байесовский классификатор*. Классический метод классификации текстов; он делает очень сильные предположения о независимости участвующих событий (появления

слов в документах), но практика показывает, что наивный байесовский классификатор оказывается весьма эффективен. Существуют две основные модели наивного классификатора. В мультиномиальной модели событием является выбор терма (слова) из множества возможных термов с вероятностями, фиксированными для данной категории, и правдоподобие документа вычисляется как:

$$p(d \mid c) = \prod_{t \in d} p(t \mid c)^{\#(t,d)},$$

где $\#(t,d)$ - количество раз, которые терм t встретился в документе d. В многомерной модели событием является попадание или непопадание терма в документ, и правдоподобие документа вычисляется как:

$$p(d \mid c) = \prod_{t \in d} p(t \mid c) \prod_{t \notin d} (1 - p(t \mid c))$$ (без учёта кратности).

2. *Метод ближайших соседей*. Этот и последующие классификаторы представляют документ в виде многомерного вектора, компоненты которого зависят от встречаемости того или иного терма в документе и других документах. В наших экспериментах использовалась классическая метрика tf-idf: значение tf-idf, соответствующее документу d и терму t, вычисляется как:

$$\text{tfidf}(t,d) = \#(t,d) \log \frac{|D|}{\#_D(t)},$$

где D – всё множество документов, $\#_D(t)$ - количество документов, в которых встречался терм t. Документ представляется как вектор длины 1 и размерности, равной размерности словаря; компоненты вектора пропорциональны значениям tf-idf. В методе ближайших соседей полученный пример классифицируется на основе взвешенной суммы желаемых категорий ближайших к нему документов.

3. *Классификатор Роккио*. В геометрическом классификаторе Роккио (Rocchio) по заданным документам сначала подсчитываются центроиды категорий с компонентами:

$$w(t,c) = \frac{\alpha}{\#_D(c)} \sum_{d \in c} w(t,d) - \frac{\beta}{\#_D(\bar{c})} \sum_{d \notin c} w(t,d),$$

а затем новый поступивший документ классифицируется в соответствии с ближайшим к нему центроидом.

4. *Метод опорных векторов*. Базовый метод опорных векторов (support vector machines, SVM) по двум данным множествам точек строит, решая задачу квадратичного программирования специального вида, оптимальную разделяющую их поверхность, т.е. такую разделяющую поверхность, расстояние от которой до ближайшей точки множеств максимально. В наших экспериментах исследовались два вида поверхностей (ядер): полиномиальное семейство и радиальные базисные функции. Для того чтобы перейти от задачи разделения двух множеств к классификации по нескольким категориям, применялся следующий метод: разделяющие поверхности обучаются для каждой категории против всех остальных, новый тестовый пример классифицируется по каждой из категорий, и в качестве окончательного результата классификации выбирается та категория, для которой расстояние до поверхности (со знаком) оказалось наибольшим.

В качестве речевой базы мы использовали базу данных SpeeCON [3] для русского языка. Эта речевая база, помимо отдельных слов, содержит записи около 550 дикторов, в свободной форме говорящих на одну из 30 заданных тем, например: «Вы звоните в банк узнать информацию по счёту» или «Вам необходимо описать текущую/недавнюю дорожную обстановку (“пробки”, дорожно-транспортные происшествия)». Записи ответов дикторов снабжены текстовками, которые использовались для обучения классификатора. Заметим, что по меркам задач классификации текстов объём обучающей выборки был весьма небольшим: в каждой из 30 категорий свободной речи базы SpeeCON мы обучались всего на ~130 высказываниях, обычно объёмом в 1-2 предложения каждое.

Автоматическое распознавание речи выполнялось на исследовательском комплексе распознавания речи, разработанном в «Центре Речевых Технологий». Основные

опубликованные результаты, относящиеся к используемому декодеру, касаются обучения контекстно-зависимых акустических моделей фонем и трифонов [4]; набор таких акустических моделей применялся и в нашем исследовании.

Схема каждого эксперимента была следующей:

- выбирается и фиксируется множество категорий;
- база речевых звукозаписей делится на обучающую и контрольную выборки;
- по текстовкам обучающей выборки составляется словарь ключевых слов;
- по этому словарю проводится распознавание файлов контрольной выборки;
- результаты распознавания подаются на вход классификатора, обученного на текстовках обучающей выборки.

Отметим, что задача автоматического порождения словаря имеет и самостоятельный интерес; порождение словаря для распознавания до сих пор было творческим процессом, и его автоматизация открывает новые возможности для таких приложений, как IVR-системы.

Таблица 1 показывает результаты работы текстовых классификаторов, запущенных на результатах распознавания речи, а также результаты распознавания во время первого прохода на этих данных. Столбцы таблицы слева направо показывают число категорий в эксперименте, референсный уровень классификации, достигнутый на истинных текстовках, без распознавания, результаты классификации мультиномиальной моделью наивного байесовского классификатора, многомерной моделью, методом ближайших соседей, методом Роккио, методом опорных векторов с полиномиальным ядром и ядром в виде гауссовских radial basis functions, число слов в словаре распознавания и два показателя успешного распознавания: доля пропущенных ключевых слов (false rejections, FR) и доля ложных срабатываний (false alarms, FA).

Таблица 1. Сравнение текстовых классификаторов

Параметры		**Словари**		**Распознавание**		**Классификация**			
# кат.	Транс-крипция	Слов	Транскрипций	FR, %	FA, %	MNB, %	NN, %	SVM, %	Nbest, %
10	вариативная	3291	16199	41,1	46,4	91,8	91,4	92,8	93,5
10	каноническая	3291	3291	35,5	55,0	94,2	94,6	94,2	96,5
30	вариативная	8819	45867	46,8	55,0	81,3	84,4	86,2	85,8
30	каноническая	8819	8819	39,6	61,8	86,1	88,1	90,3	89,2

Оказывается, что, несмотря на крайне низкое качество распознавания, классификаторам, тем не менее, удаётся получить на таких данных хорошие результаты: около 95% на 10 категориях и около 90% - на всех 30 категориях. В таблице также приведены, для сравнения, результаты кросс-валидационного тестирования на текстовках (т.е. классификация не результатов распознавания, а истинных текстовок). Видно, что для малого числа категорий качество «зашумленной» классификации приближается к референсному.

Интересной замеченной нами тенденцией (не вошедшей в таблицу 1) стало то, что с ростом объёма словаря (для одного и того же набора категорий), который регулировался порогом прироста информации, качество распознавания, естественно, падало, но качество последующей классификации зачастую росло, вероятно, за счёт того, что распознавалось больше термов, в том числе больше «правильных» термов. Вообще, влияние доли пропущенных ключевых слов (FR) на качество классификации значительно выше, чем влияние ложных срабатываний (FA).

3. Двухпроходное распознавание

Общая архитектура предлагаемого подхода к многопроходному распознаванию проста и уже была предложена как нами [5], так и другими исследователями в контексте других языков [6]. Сначала поступающий на вход речевой сигнал распознается с общим большим

словарём, обученным по всей тестовой базе. Затем распознанный текст классифицируется тем или иным текстовым классификатором (см. раздел 2) и распознаётся снова с уточнёнными грамматикой и словарём до тех пор, пока не выполнится условие выхода из цикла. Подход, естественно, может быть применён к «дереву грамматик» любой глубины, но в приведённых экспериментах (и большинстве практических применений) достаточно рассматривать «дерево» глубины два: во время первого прохода речь распознаётся с «общим» словарём, затем классифицируется, а затем, во время второго прохода, распознаётся с «индивидуальным» словарём, основанным только на той категории, к которой был отнесёт распознанный на первом проходе текст.

Обнадёживающие результаты классификации позволяли надеяться на то, что в результате двухпроходного распознавания будут достигнуты значительное сокращение ошибок. И действительно, из таблицы 2, где представлены результаты двухпроходного распознавания, видно, что результаты распознавания после классификации стали на порядок лучше, чем были до неё. Проанализируем эти результаты более подробно.

В таблице 2 представлены:

- референсные показатели качества распознавания для ситуации, когда каждый файл распознаётся сразу со словарём правильной категории, без фазы классификации (столбцы «Индивидуальные словари»); результаты даны для двух вариантов индивидуальных словарей разного размера;
- результаты первого прохода распознавания (первая строка каждого раздела «Второй проход»);
- результаты второго прохода распознавания (раздел «Второй проход») для тех же вариантов индивидуальных словарей.

Таблица 2. Результаты второго прохода распознавания

Параметры			Индивидуальные словари		Второй проход	
Число категорий	Транс-крипция	Словари, слов	FR, %	FA, %	FR, %	FA, %
10	вариативная	---	первый проход:		41,1	46,4
		100-300	32,4	22,4	33,2	23,0
		150-600	33,3	24,4	34,5	24,3
10	каноническая	---	первый проход:		35,5	55,0
		100-300	35,0	20,1	36,0	20,5
		150-600	38,8	27,1	40,0	27,9
30	вариативная	---	первый проход:		46,8	55,0
		100-300	32,9	24,7	35,3	25,6
		150-600	25,2	25,9	35,4	25,5
30	каноническая	---	первый проход:		39,6	61,8
		100-300	35,7	21,3	37,4	21,8
		150-600	29,3	28,6	31,0	29,5

Видно, что при переходе от первого прохода распознавания ко второму значительно снижаются обе доли ошибок, но особенно значительный выигрыш достигается по отношению к доле ложных срабатываний. Результаты двупроходного распознавания чрезвычайно близки к «идеальным» результатам распознавания по индивидуальным словарям, значительно ближе, чем можно было предположить из результатов классификации, ведь распознавание с неправильным индивидуальным словарём неизбежно должно порождать массу ошибок. По нашему мнению, это происходит потому, что для того чтобы распознанный файл неверно классифицировался, необходимо, чтобы распознавание было совсем плохим, настолько, что даже повторное распознавание с неподходящим словарём не может его серьёзно ухудшить.

4. Заключение

В настоящей работе мы показали, что методика двухпроходного распознавания с уточнением словаря на основе классификации приводит к значительному улучшению качества распознавания речи.

Предложенную методику лучше всего применять для ситуаций, в которых распознаваемая речь естественным образом может быть классифицирована на несколько тематических категорий. В качестве примеров таких применений можно привести IVR-системы (абонент, позвонив в автоматическую систему, обычно хочет воспользоваться одной из заранее определённых функций системы), системы мониторинга и анализа данных колл-центров (классификация входящих звонков и последующий статистический анализ), мониторинг радио- и телепрограмм.

Литература

1. Николенко С.И., Кореневский М.Л., Пономарёва И.А., Левин К.Е. Классификация звуковых файлов посредством классификации результатов распознавания. Труды V Международной конференции по прикладной лингвистике "Прикладная лингвистика в науке и образовании", Санкт-Петербург, РГПУ, 27-28 марта 2010 г.
2. Sebastiani F. Machine learning in automated text categorization // ACM Computing Surveys. 2002. Vol. 34, no. 1, pp. 1-47.
3. Siemund R., Hoge H., Kunzmann S., Marasek K. SPEECON - speech data for consumer devices // Proceedings of the Second International Conference on Language Resources and Evaluation. Vol. II. 2000, pp. 883-886.
4. Tatarnikova M., Tampel I., Oparin I., Khokhlov Y. Building Acoustic Models for Large Vocabulary Continuous Speech Recognizer for Russian // Proceedings of the 11 th International Conference on Speech and Computer, 2006.
5. Николенко С. И., Левин К. Е., Хохлов Ю. Ю. Двухпроходное автоматическое распознавание речи с использованием интеллектуального анализа текстов // Труды конференции ИММВИИ-2009, т. 1, М., Физматлит, 2009, С. 192-202.
6. Lane I.R., Kawahara T., Matsui T., Nakamura S. Dialogue Speech Recognition by Combining Hierarchical Topic Classification and Language Model Switching // IEICE - Transactions on Information and Systems, vol. E88-D, issue 3, 2005, pp. 446-454.

И.С. Кипяткова

Исследование статистических *n*-граммных моделей языка для распознавания слитной русской речи со сверхбольшим словарем

Санкт-Петербургский институт информатики и автоматизации РАН,
Санкт-Петербург, Россия,
kipyatkova@iias.spb.su

1. Введение

Одной из основных нерешенных на данный момент проблем в области распознавания речи является автоматическое стенографирование или распознавание слитной разговорной речи со сверхбольшим словарем (сотни тысяч и миллионы слов). В этой задаче для генерации грамматически правильных и осмысленных гипотез произнесенной фразы распознавателю речи необходима некоторая модель языка или грамматика, описывающая допустимые фразы. Процесс распознавания речи может быть представлен как поиск наиболее вероятной последовательности слов [1]:

$$W = \arg\max_{W} P(W \mid A) = \arg\max_{W} P(A \mid W) \cdot P(W),$$

где $P(A|W)$ - это вероятность появления гипотезы по оценке акустической модели и $P(W)$ - это вероятность появления гипотезы по оценке языковой модели.

Для многих языков (например, для английского) разработаны методы создания моделей языка, которые позволяют повысить точность распознавания речи. Но эти методы не могут быть напрямую применены для русского языка из-за свободного порядка слов в предложениях и наличия большого количества словоформ для каждого слова.

Одной из наиболее эффективных моделей естественного языка является статистическая модель на основе *n*-грамм слов, цель которой состоит в оценке вероятности появления цепочки слов $W=(w_1,w_2,\ldots w_m)$ в некотором тексте. *n*-граммы представляют собой последовательность из *n* элементов (например, слов), а *n*-граммная модель языка используется для предсказания элемента в последовательности, содержащей *n*-1 предшественников. Эта модель основана на предположении, что вероятность какой-то определенной *n*-граммы, содержащейся в неизвестном тексте, можно оценить, зная, как часто она встречается в некотором обучающем тексте.

Вероятность $P(w_1,w_2,\ldots,w_m)$ можно представить в виде произведения условных вероятностей входящих в нее *n*-грамм [2]:

$$P(w_1, w_2, \ldots, w_m) = \prod_{i=1}^{m} P(w_i \mid w_{1,} w_{2,} \ldots w_{i-1}),$$

или аппроксимируя $P(W)$ при ограниченном контексте длиной *n*-1:

$$P(w_1, w_2, \ldots, w_m) \cong \prod_{i=1}^{m} P(w_i \mid w_{i-n+1,} w_{i-n+2,} \ldots w_{i-1}).$$

Вероятность появления *n*-граммы вычисляется на практике следующим образом:

$$P(w_i \mid w_{i-n+1,} \ldots w_{i-1}) = \frac{C(w_{i-n+1,} \ldots w_i)}{C(w_{i-n+1,} \ldots w_{i-1})},$$

где C – количество появлений последовательности в обучающем корпусе.

В разделе 2 описывается процесс сбора и предварительной обработки текстовых данных для создания статистической модели русского языка. В разделе 3 приводится статистический анализ собранного корпуса, делается проверка соответствия корпуса закону Ципфа, даются результаты по подсчету частоты появления различных *n*-грамм слов, описываются созданные биграммная и триграммная модели русского языка. В разделе 4

приводятся результаты распознавания тестового корпуса русской речи с использованием созданных моделей языка.

2. Сбор и автоматическая обработка текстовых материалов

Существуют несколько текстовых корпусов русского языка, например, "Национальный корпус русского языка" (www.ruscorpora.ru) и "Корпус русского литературного языка" (www.narusco.ru), которые содержат, в основном, текстовый материал конца XX века. Эти корпуса включают в себя различные типы текстов: художественный, публицистический, научный, а также содержат в небольшом объеме и стенограммы устной речи. В работе [3] описан новостной корпус, собранный из примерно двух тысяч СМИ-источников. Объем этого корпуса составляет 7,3 млрд словоупотреблений.

Нами для создания модели языка был собран и обработан новостной текстовый русскоязычный корпус, сформированный из новостных лент последних лет четырех интернет-сайтов: www.ng.ru ("Независимая газета"), www.smi.ru ("СМИ.ru"), www.lenta.ru ("LENTA.ru"), www.gazeta.ru ("Газета.ru"). Он содержит тексты, отражающие срез современного состояния языка, в том числе и разговорного русского языка. Пополнение этого корпуса может осуществляться автоматически при обновлении сайтов в режиме он-лайн, что позволяет оперативно добавлять новые появляющиеся в языке слова и переобучать модель языка с учетом новых текстовых данных. Естественный язык, будучи открытой системой, постоянно изменяется с изменением общественной жизни, развитием новых областей знаний, и он-лайн пополнение текстового корпуса позволяет учитывать изменения, происходящие в языке.

Автоматическая обработка собранного материала осуществляется следующим образом. Вначале происходит разбиение текстового массива на предложения, которые должны начинаться либо с заглавной буквы, либо с цифры. При этом учитывается, что в начале предложения могут стоять кавычки. Предложение заканчивается точкой, восклицательным или вопросительным знаком либо многоточием. Кроме того, при разделении текста на предложения учитывается, что внутри предложения могут стоять инициалы и/или фамилии. Формально это похоже на границу раздела двух предложений, поэтому если точка идет после одиночной заглавной буквы, то эта точка не будет считаться концом предложения. Предложения, содержащие прямую и косвенную речь, разделяются на отдельные предложения. При этом возможны три случая:

а) прямая речь идет после косвенной;
б) прямая речь идет до косвенной;
в) косвенная речь находится внутри прямой речи.

В первом случае формальными признаками, при которых происходит выделение прямой и косвенной речи, является наличие двоеточия, после которого следуют кавычки. Во втором случае разделение происходит, если после кавычек стоит запятая, а затем тире. В третьем случае исходное предложение разбивается на три предложения: первое – от кавычек до запятой и тире, второе – то, что находится между первой запятой с тире до второй запятой с тире, третье – от запятой с тире до конца предложения. Начало и конец предложения отмечаются знаками <s> и <\s> соответственно. После разделения текстового материала на предложения выполняется его нормализация. Происходит удаление текста, написанного в любых скобках, удаление предложений, состоящих из пяти и меньшего количества слов (как правило - это заголовки, составленные не по грамматическим правилам для полных предложений). Затем из текстов удаляются знаки препинания, расшифровываются общепринятые сокращения (например, "см", "кг"). В словах, начинающихся с заглавной буквы, происходит замена заглавной буквы на строчную. Если все слово написано заглавными буквами, то замена не делается, так как это слово, вероятно, является аббревиатурой. На данный момент общий объем корпуса после его обработки составляет свыше 200 млн словоупотреблений (около 1 Гб данных).

3. Статистическая обработка текстового корпуса и создание модели языка

На базе собранного русскоязычного текстового корпуса был создан частотный словарь, размер которого составляет около одного миллиона уникальных словоформ, а также для данного корпуса определена частота встречаемости различных биграмм и триграмм. Была выполнена проверка соответствия текстового корпуса закону Ципфа (см. рисунок 1). Известно, что закон Ципфа — эмпирическая закономерность распределения частоты слов естественного языка: если все слова языка в достаточно большом осмысленном тексте упорядочить по убыванию частоты их использования, то частота слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру [4]. Теоретически, для текстового корпуса график такой зависимости должен иметь вид, как показано на рисунке 1 пунктирной линией. Полученный экспериментально график имеет незначительное отклонение от этой линии, но в целом можно заключить, что собранный корпус соответствует закону Ципфа.

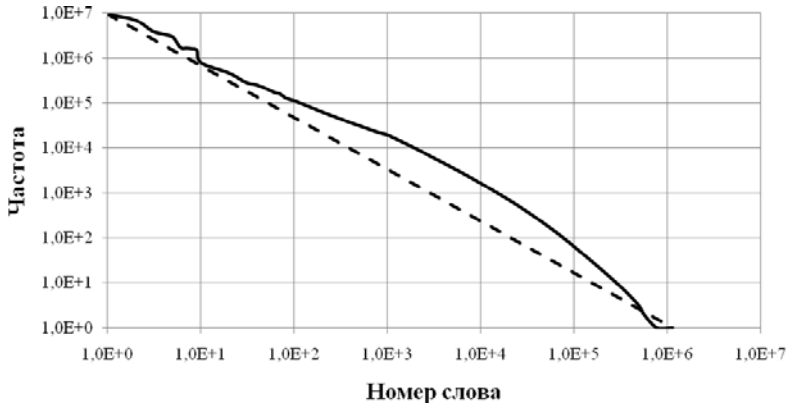
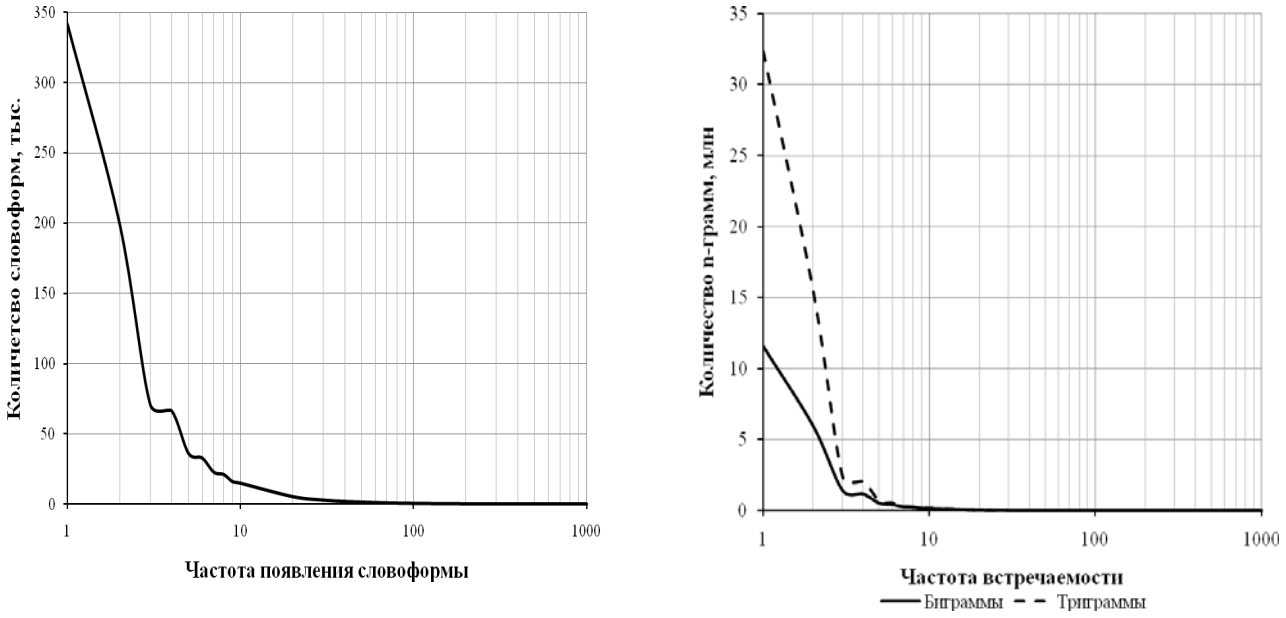


Рис. 1. Проверка соответствия текстового корпуса закону Ципфа.

Для автоматического распознавания речи необходимо иметь словарь фонематических транскрипций слов. Транскрипции для слов из собранного корпуса в основном создавались с помощью программного модуля, позволяющего создавать фонематические транскрипции слов автоматически [5]. Для генерации транскрипций модулю необходима база данных словоформ русского языка с отметкой ударения. В качестве таковой использовалась база данных словоформ русского языка, созданная путем объединения двух баз данных, свободно доступных в Интернете: 1) морфологическая база данных проекта STARLING (starling.rinet.ru); 2) морфологическая база данных проекта АОТ (www.aot.ru). Первая база данных содержит около 1 млн 800 тыс. различных словоформ, это количество является недостаточным для наших исследований. В этой базе данных для некоторых сложных слов проставлено также второстепенное ударение. Вторая база данных содержит свыше 2 млн 200 тыс. словоформ. Однако в этой базе данных, в отличие от первой, отсутствует буква *ё* и информация о второстепенном ударении. Поэтому эти две базы данных были объединены. Кроме того, в этот словарь было добавлено более 40 тыс. словоформ из обрабатываемого тестового материала. Объем получившегося словаря превысил 2 млн 380 тыс. различных словоформ. Вручную были сделаны транскрипции для аббревиатур, а также для некоторых широко распространенных в наше время слов, заимствованных из английского языка.

На рисунке 2 представлены графики распределения частот встречаемости униграмм (рисунок 2а), биграмм и триграмм (рисунок 2б). Эти графики показывают, что в текстах присутствует достаточно большое число редких слов. Более 350 тыс. слов встретились в текстовом корпусе только один раз. Часто, это слова, написанные с опечатками. Поэтому для сокращения списка *n*-грамм и скорости обработки целесообразно удалять редкие *n*-граммы. Для этого был введен порог К, *n*-граммы, которые встретились меньше К раз, удаляются из списка. На рисунке 3 показано, как уменьшается количество *n*-грамм с ростом К. При удалении из списка *n*-грамм, которые встретились только 1 раз (К=2), список *n*-грамм

сократился вдвое, а при К=3 список сократился еще в несколько раз. При дальнейшем увеличении К сокращение было уже незначительным.



а) б)

Рис. 2. Распределение частоты встречаемости различных *n*-грамм слов (а – униграмм; б – биграмм и триграмм).

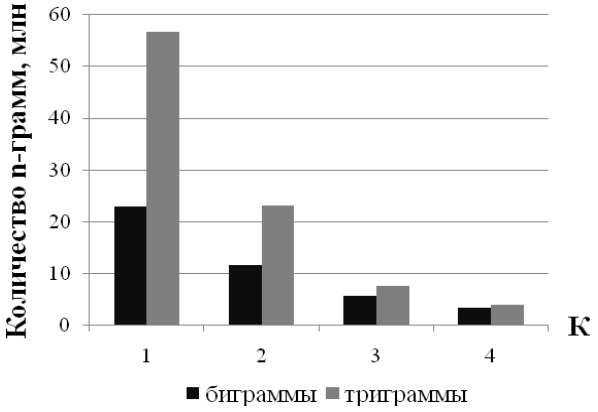


Рис. 3. Распределение количества *n*-грамм при минимальном пороге встречаемости.

Статистическая модель языка была создана с помощью программного модуля обработки и анализа текстов CMU (Cambridge Statistical Language Modeling Toolkit) [6]. Модель языка создавалась в несколько этапов. Вначале количество биграмм составляло 22,85 млн, триграмм - 56,70 млн, количество уникальных слов (словарь) – 937 тыс., то есть это сверхбольшой словарь [7]. Далее для сокращения количества редких слов при построении модели языка был введен порог *K*=4, то есть *n*-граммы, у которых частота появления меньше 4, удалялись из модели языка. Затем для слов, которые использовались в этих моделях языка, были автоматически созданы транскрипции. *n*-граммы со словами, для которых транскрипции не могли быть созданы автоматически (поскольку этих слов не было ни в словаре ударений, ни в списке аббревиатур, ни в списке иностранных слов), были удалены из модели языка. Однако из-за удаления некоторых *n*-грамм из модели языка появилось много слов, которые в модели не приводят к конечному результату (разрывают цепочку слов), поскольку встречаются в *n*-граммах не во всех позициях. Поэтому модель языка была также сокращена путем удаления *n*-грамм, содержащих такие слова. В результате в конечной биграммной модели количество уникальных словоформ составило 182 тыс.,

количество биграмм – 3,17 млн, в триграммной модели количество уникальных словоформ – 126 тыс., триграмм – 3,54 млн. В таблице 1 представлено количество *n*-грамм, которое получалось на этих этапах.

Таблица 1. Размер словаря и количество n-грамм на этапах создания модели языка

Этап создания модели языка	Биграммная модель		Триграммная модель	
	Размер словаря, тыс. слов	Количество *n*-грамм, млн	Размер словаря, тыс. слов	Количество *n*-грамм, млн
До удаления n-грамм	937	22,85	937	56,70
После удаления n-грамм с порогом К=4	273	3,35	201	3,90
После удаления n-грамм со словами без транскрипций	234	3,23	176	3,64
Конечная модель	182	3,17	126	3,54

4. Результаты распознавания речи с использованием модели языка

Для распознавания слитной русской речи использовался разработанный в СПИИРАН декодер SIRIUS [8]. Запись обучающего и тестового речевого корпуса для системы производилась с частотой дискретизации 44 кГц, 16 бит на отсчет, моно, отношение сигнал/шум больше 35дБ. Для обучения системы распознавания речи были использованы около 300 фраз из речевого корпуса [9]. Система была обучена на записях одного диктора и является, таким образом, дикторозависимой. Для тестирования системы использовались 516 слитно произнесенных фраз, состоящих из 3667 слов, длина каждой фразы не превышала 20 слов. При этом 100 фраз были выбраны из описанного выше новостного корпуса, еще 100 фраз взяты из материалов интернет-газеты "Фонтанка.ru" (www.fontanka.ru), остальные 316 фраз взяты из речевого корпуса [9]. В тестовом корпусе количество биграмм, присутствующих в модели языка, составляет 77,47 %, триграмм – 31,49 %. Для этого тестового корпуса вычислены величины энтропии и коэффициента неопределенности (perplexity) статистической модели языка. По определению, информационная энтропия — мера хаотичности информации, неопределенность появления какого-либо символа первичного алфавита. При отсутствии информационных потерь она численно равна количеству информации на символ передаваемого сообщения. Поскольку тексты на естественном языке могут рассматриваться в качестве информационного источника, энтропия вычисляется по следующей формуле [10]:

$$H = -\lim_{m\to\infty}\frac{1}{m}\sum_{w_1,w_2,\ldots,w_m}\left(P(w_1,w_2,\ldots,w_m)\log_2 P(w_1,w_2,\ldots,w_m)\right).$$

Это суммирование делается по всем возможным последовательностям слов. Но поскольку язык является эргодичным источником информации [10], выражение для вычисления энтропии будет выглядеть следующим образом:

$$\hat{H} = -\frac{1}{m}\log_2 P(w_1,w_2,\ldots,w_m).$$

Коэффициент неопределенности является параметром, по которому оценивается качество *n*-граммных моделей языка, и вычисляется следующим образом [10]:

$$PP = 2^{\hat{H}} = \hat{P}(w_1,w_2,\ldots,w_m)^{-\frac{1}{m}},$$

где $\hat{P}(w_1,w_2,\ldots,w_m)$ - это вероятность появления последовательности слов $w_1,w_2,\ldots,w_m$. Коэффициент неопределенности показывает среднее значение наиболее вероятных слов, которые могут следовать за данным словом. Для униграммной модели величина коэффициента неопределенности составила 6109,24, энтропии – 12,58 бит/слово, для биграммной модели коэффициент неопределенности равен 1325,18, энтропия –

10,37 бит/слово, для триграммной модели коэффициент неопределенности равен 890,84, энтропия – 9,80 бит/слово.

При оценке качества распознавания речи или текста выделяют три типа ошибок: замена одного слова другим, удаление слова, вставка слова [11]. Существует два критерия, по которым определяется качество распознавания: правильность и точность распознавания. Правильность определяется следующим образом [11]:

$$C = \frac{H}{N} \cdot 100\%,$$

где H – количество правильно распознанных слов, N – общее число слов в распознаваемом сообщении (в данного тестового корпуса N=3667).

$$H = N - D - S,$$

где D и S – количество неверно удаленных и замененных слов соответственно. Этот критерий не учитывает ошибки, связанные со вставкой слов, поэтому обычно используется другой критерий – точность распознавания, который вычисляется по следующей формуле [11]:

$$A = \frac{H - I}{N} \cdot 100\%,$$

где I – количество неверно вставленных слов.

В качестве фонетических единиц при распознавании речи применялись контекстно-зависимые фоны (трифоны). Для 0-граммной, униграммной и биграммной модели использовался одинаковый словарь, объемом в 182 тыс. слов, для триграммной модели объем словаря был 126 тыс. слов. Результаты распознавания с применением различных моделей языка представлены в таблице 2. При использовании 0-граммной модели, то есть при распознавании без вероятностной модели [12], точность распознавания слов оказалась отрицательной и равной почти -100 % (ошибок было больше, чем слов во входной фразе). При этом большую часть ошибок составили ошибочно вставленные слова. В основном вставлялись короткие слова, или же длинное слово при распознавании разбивалось на несколько коротких, например, слово "стали" было распознано как "с таль и". Использование биграммной модели позволило повысить точность распознавания слов до 11,04 % и значительно сократить число вставленных слов. При использовании этой модели большую часть ошибок составили замены слов, однако доля ошибок, связанных со вставкой слов, тоже достаточно велика. Значительная часть коротких слов – это предлоги, союзы и частицы, у которых частота появления очень большая.

Таблица 2. Результаты распознавания слов с использованием различных моделей языка

Тип модели	Размер словаря (V), тыс. слов	Вероятность появление i-го слова	A, %	C, %	H	D	S	I
0-граммная	182	$P(w) = 1/\|V\|$	-99,15	21,95	805	32	2830	4441
1-граммная	182	$P(w)$	33,38	43,11	1581	303	1783	357
2-граммая	182	$P(w_i\|w_{i-1})$	11,04	53,07	1946	75	1646	1541
2-граммная модифицированная	182	$P(w_i\|w_{i-1})$	25,52	45,13	1655	147	1865	719
3-граммная	126	$P(w_i\|w_{i-2},w_{i-1})$	24,71	25,63	940	1124	1603	34

Для того чтобы сократить количество вставок, был составлен список однофонемных и двухфонемных предлогов, союзов и частиц, и вероятность биграмм с такими словами была уменьшена до 10^{-10}. Эта модифицированная модель позволила сократить количество вставок, и, хотя число удалений и замен слов немного увеличилось, точность распознавания повысилась до 25,52 %. При использовании триграммной модели точность распознавания равна 24,71 %. Стало довольно неожиданным, что наилучшие результаты достигнуты с

использованием униграммной модели, где точность распознавания составила 33,38 %. Это связано с тем, что достаточно много биграмм и триграмм из тестового корпуса отсутствуют в обучающем материале.

В таблице 3 представлены результаты распознавания слов с использованием нескольких разновидностей униграммной модели. Модифицированная униграммная модель была создана аналогично модифицированной биграммной модели, то есть фонетически коротким словам из списка приписывалась вероятность 10^{-10}. Это позволило повысить точность распознавания до 35,86 %. Затем словарь для униграммной модели был расширен. В него были включены все словоформы из новостного корпуса, для которых возможно создать транскрипции автоматически. Объем этого словаря составил 535 тыс. словоформ. Точность распознавания при применении униграммной модели языка с расширенным словарем составила 35,59 %, то есть немного выше, чем с униграммной моделью в 182 тыс. словоформ. Наилучшая точность распознавания (36,92 %) получилась при использовании модифицированной униграммной модели со словарем 535 тыс. слов. Для униграммных моделей основную долю ошибок составили неправильные замены слов, при этом их количество для всех разновидностей этой модели приблизительно одинаково. Улучшение в качестве распознавания в основном связано с уменьшением числа вставок слов при распознавании.

Таблица 3. Результаты распознавания с различными вариантами униграммных моделей

Тип модели	Размер словаря (V), тыс. слов	A, %	C, %	H	D	S	I
1-граммная	182	33,38	43,11	1581	303	1783	357
1-граммная модифицированная	182	35,86	38,51	1412	463	1792	97
1-граммная с расширенным словарем	535	35,59	43,63	1600	316	1751	295
1-граммная модифицированная с расширенным словарем	535	36,92	39,00	1430	488	1749	76

5. Заключение

Текстовые данные для создания модели языка были собраны из материалов интернет-сайтов четырех электронных газет. Таким образом, собранный корпус основывается на текстах с большим количеством стенограмм выступлений и прямой речи, отражающих особенности современного языка, а не на литературных текстах, которые крайне далеки от разговорной речи. Разработанная методика сбора текстового материала позволяет при обновлении интернет-сайтов оперативно дополнять текстовый корпус и затем переобучать модель языка в режиме он-лайн, учитывая тем самым изменения, происходящие как в самом языке, так и в контексте текущих событий. Однако использование интернет-материалов имеет и ряд недостатков, главным из которых является наличие в текстах опечаток. Кроме того, в таких текстах присутствует много имен собственных, большинство из которых в разговорной речи встречается редко. Из-за этого возрастает объем созданных в результате обработки текста n-грамм.

На основе собранного корпуса были созданы униграммная, биграммная и триграммная модели русского языка. В ходе экспериментов по распознаванию слитно произнесенных фраз оказалось, что униграммная модель эффективнее, чем биграммная и триграммная модели, если сравнивать точность распознавания слов в слитной речи. Это связано с тем, что в тестовом материале было много биграмм и триграмм, которые отсутствуют в обучающем корпусе. Решить данную проблему, возможно, поможет использование метода возврата (backing-off method) [10]. Суть данного метода заключается в том, что если некоторая n-грамма отсутствует в обучающем корпусе или частота ее появления очень низкая, тогда вместо нее используется вероятность (n-1)-граммы.

Данное исследование поддержано Минобрнауки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» (госконтракты № П2579 и П2360), фондом РФФИ (проекты № 08-08-00128 и 09-07-91220-СТ), а также Фондом "Научный потенциал" (договор № 201).

Литература

1. Rabiner L., Juang B.-H. Fundamentals of Speech Recognition. Prentice Hall, 1995, 507 p.
2. Benesty J., Sondhi M., Huang Y. (eds.) Springer Handbook of Speech Processing. Springer, 2008, 1176 p.
3. Баглей С.Г., Антонов А.В., Мешков В.С., Суханов А.В. Статистические распределения слов в русскоязычной текстовой коллекции. Материалы международной конференции "Диалог 2009". Москва, 2009, С. 13-18.
4. Gelbukh A., Sidorov G. Zipf and Heaps Laws' Coefficients Depend on Language. Proc. CICLing-2001, Conference on Intelligent Text Processing and Computational Linguistics, 2001, Mexico City. Lecture Notes in Computer Science № 2004, Springer-Verlag, pp. 332–335.
5. Кипяткова И.С., Карпов А.А. Разработка и оценивание модуля транскрибирования для распознавания и синтеза русской речи // Искусственный интеллект, Донецк, Украина, №3, 2009, С. 178-185.
6. Clarkson P., Rosenfeld R. Statistical language modeling using the CMU-Cambridge toolkit, in Proc. EUROSPEECH, Rhodes, Greece, 1997, pp. 2707–2710.
7. Hori, T., Nakamura A. An extremely-large-vocabulary approach to named entity extraction from speech. Proceedings of ICASSP'2006, Toulouse, France, 2006.
8. Ronzhin A., Leontieva An., Kagirov I., Karpov A., "Morpho-Phonetic Tree Decoder for Russian", In Proc. of 12-th International Conference on Speech and Computer SPECOM, Moscow, Russia, 2007, pp. 491-498.
9. Jokisch O., Wagner A., Sabo R., Jaeckel R., Cylwik N., Rusko M., Ronzhin A., Hoffmann R. Multilingual Speech Data Collection for the Assessment of Pronunciation and Prosody in a Language Learning System. In Proc. 13-th International Conference SPECOM'2009, St. Petersburg, 2009, pp. 515-520.
10. Moore G.L. Adaptive Statistical Class-based Language Modelling. PhD thesis, Cambridge University, 2001, 193 p.
11. Young S. et al. The HTK Book (for HTK Version 3.4). Cambridge, UK, 2009.
12. Merkel A., Klakow D. Improved Methods for Language Model Based Question Classification. Proceedings of 8th Interspeech Conference, Antwerp, 2007.

Ю.Г. Бондарос

Программно-аппаратный комплекс ”ТРЕНЕР” для тестирования систем распознавания речевых команд[1]

ФГУП ГосНИИ АС,
Москва, Россия,
bondaros@gosniias.ru

1. Введение

Для речевых интерфейсов систем с высокими требованиями по обеспечению точности распознавания команд в технологическом процессе разработки и сертификации программного обеспечения большое место по трудоемкости занимает процесс тестирования. При разработке авиационных приложений тестирование производится многократно и его проведение возможно в условиях лаборатории, ангара, кабины пилота самолета на стоянке или в условиях реального полета. При этом система распознавания команд может быть интегрирована в сетевое программное обеспечение кабины пилота или сопряжена со специально разработанным программно-аппаратным комплексом для тестирования. С целью снижения затрат на разработку максимально подробное исследование приходится проводить в условиях лаборатории. Некоторые эксперименты по оценке точности распознавания в достаточно полном объеме можно провести практически только в условиях лаборатории. Эксперимент необходимо проводить в условиях акустико-фоновой обстановки, максимально приближенной к условиям кабины пилота. При этом желательно, чтобы методика эксперимента обеспечивала ограниченное пребывание оператора во вредных условиях повышенного уровня шума.

Исследования по оценке точности распознавания предполагают изучение зависимости этой точности от:

- голоса диктора;
- манеры диктования (состояния диктора);
- громкости речи диктора;
- темпа речи диктора;
- характеристик микрофонного устройства;
- положения микрофонного устройства в пространстве;
- акустико-фоновой обстановки (этапа полёта).

Результатом тестирования системы распознавания команд служат следующие параметры:

- доля распознанных команд *IRR (Instruction Recognition Rate):*

$$IRR = N_R / M_R \quad ;$$

- доля неверно распознанных команд *IER (Instruction Error Rate):*

$$IER = N_E / M_R \quad ;$$

- доля нераспознанных команд *IMR (Instruction Miss Rate):*

$$IMR = N_M / M_R \quad ,$$

где N_R - число верно распознанных команд, N_E - число неверно распознанных команд, N_M - число нераспознанных команд, M_R - число верно произнесенных диктором команд, M_E - число неверно произнесенных команд.

[1] Работа выполнена при поддержке РФФИ. Грант РФФИ 09-08-00763-а

Дополнительно оцениваются:

- доля распознанных слов *WRR (Word Recognition Rate):*

$$WRR = 1 - WER; \quad WER = \frac{S + D + I}{N};$$

- коэффициент готовности диктора *SRR (Speaker Readiness Ratio),* или доля верно произнесенных диктором команд:

$$SRR = M_R / (M_R + M_E) \;;$$

- время реакции системы *SRT (System Reaction Time):*

$$SRT = t_s - t_i,$$

где S - число слов, которые надо заменить, D - число слов, которые надо уничтожить, I - число слов, которые надо вставить в полученный после распознавания текст, чтобы получить без ошибок текст теста, содержащего *N* слов, t_s - время появления стимула для произнесения команды, t_i - время появления кода распознанной команды.

Ниже описан программно-аппаратный комплекс для тестирования систем распознавания речевых команд в условиях лаборатории.

2. Описание программно-аппаратный комплекса ”ТРЕНЕР”

В состав программно-аппаратного комплекса для тестирования систем распознавания речевых команд в условиях лаборатории входят:

I. Компьютер с установленным в нем программным обеспечением “Тренер”, интегрированным с испытуемой системой распознавания речевых команд и движками синтезаторов русской и английской речи.

II. Микрофон или гарнитура

III. Установка “Рот искусственный” (или “Голова и торс”).

IV. Имитатор акустико-фоновой обстановки (компьютер с аудиосистемой для создания акустико-фоновой обстановки).

Комплекс позволяет по выбору выполнять тестирование в одном из четырех режимов:

1. Непосредственное диктование речевых команд диктором,

2. Воспроизведение через “Рот искусственный” речевых команд по предварительно подготовленным аудиозаписям команд,

3. Синтез команд по предварительно подготовленным аудиозаписям слов и воспроизведение синтезированных речевых команд через “Рот искусственный”,

4. Воспроизведение через “Рот искусственный” аудиосигналов, создаваемых с помощью синтезатора речи.

Схема комплекса в режиме диктования речевых команд диктором представлена на Рис. 1. Предварительно (до начала диктования) устанавливается (выбирается из перечня тестов) набор команд для тестирования (тест), и устанавливается темп диктования. Текст команды периодически, в соответствии с установленным темпом, появляется на экране компьютера как подсказка диктору. При неверном произнесение диктором команды результаты распознавания по команде оператора или диктора (нажатием клавиши) исключаются из данных для оценки точности системы распознавания команд, но учитываются при оценке готовности диктора.

Рис. 1 Схема комплекса “Тренер” в режиме диктования диктора.

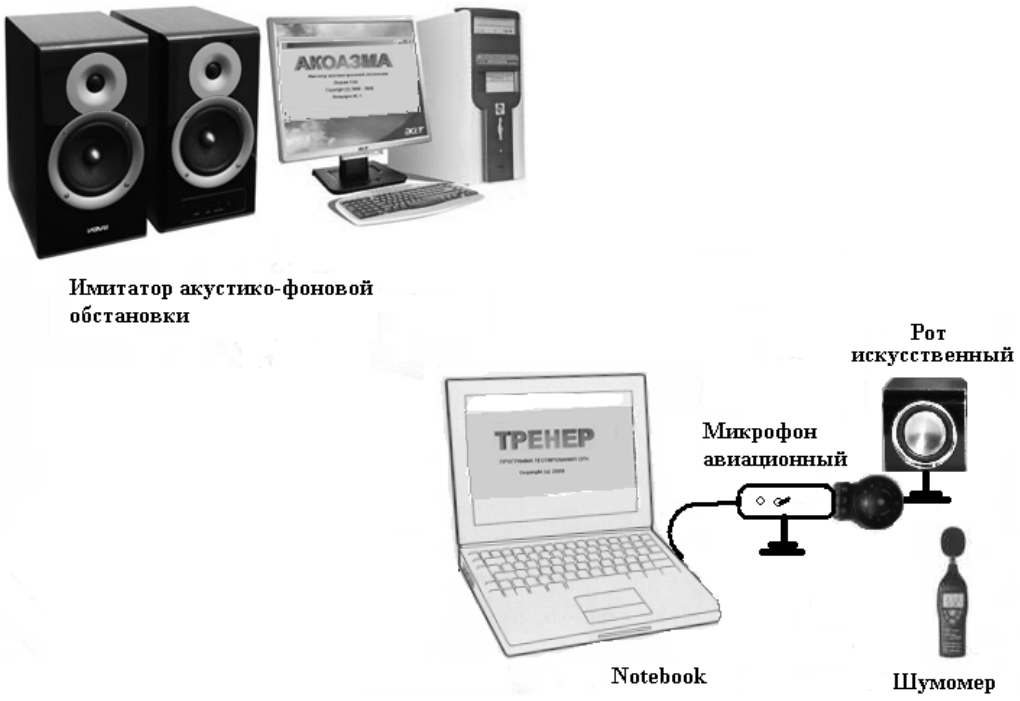


Рис. 2 Схема комплекса “Тренер” в режиме воспроизведения аудиофайлов, синтеза команд по предварительно подготовленным аудиозаписям слов или использования синтезатора.

Режимы тестирования путем воспроизведения аудиофайлов синтеза команд по предварительно подготовленным аудиозаписям слов или использования синтезатора являются альтернативными. Схема комплекса в этих режимах представлена на Рис. 2. В режиме воспроизведения аудиофайлов или синтеза команд по предварительно подготовленным аудиозаписям слов выполняются установки, аналогичные установкам в режиме диктования диктором, и дополнительно устанавливается громкость воспроизведения. В режиме диктования с помощью синтезатора дополнительно установкам,

используемым в режиме диктования диктором, устанавливается частота основного тона, скорость речи и её громкость. На рисунке 3 показан пример диалогового окна программы.

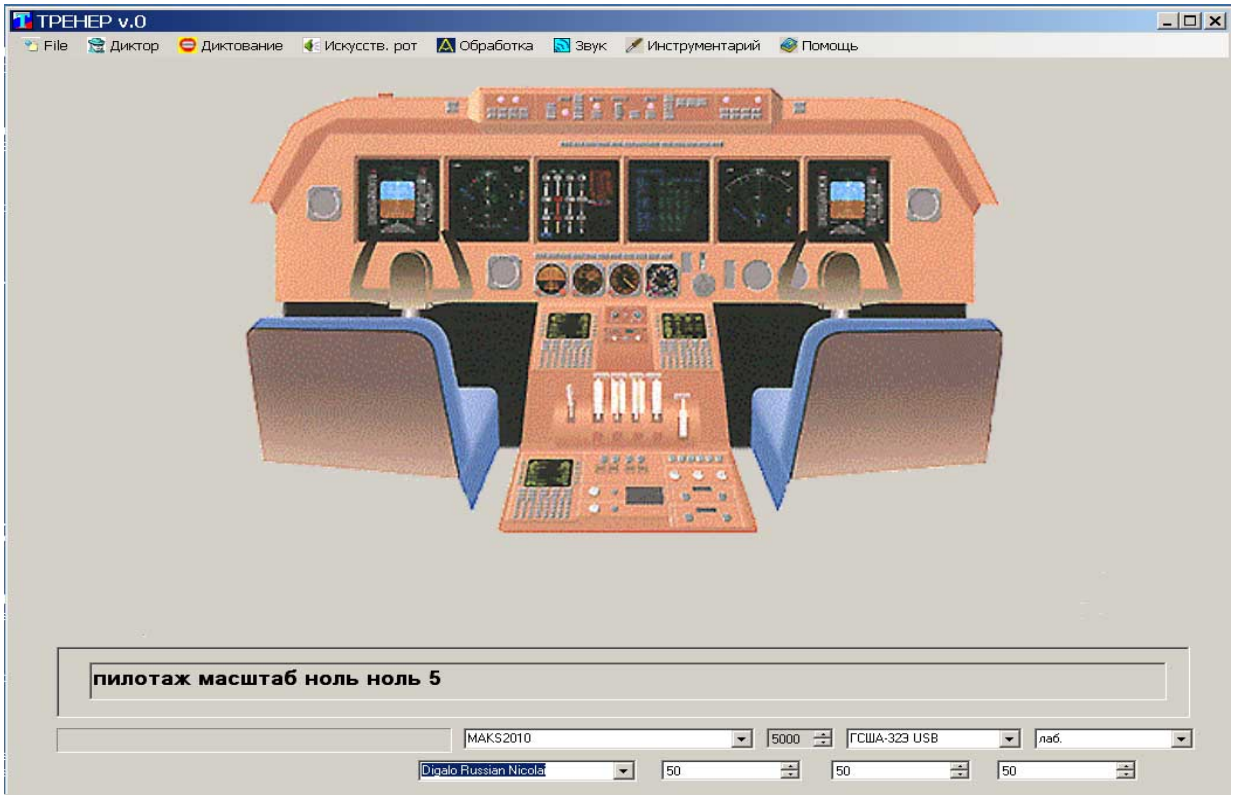


Рис. 3. Основное меню программы “Тренер”.

При разработке комплекса учтены международные и российские рекомендации [1-9]. Изменение звукового давления, развиваемого динамиком установки “Рот искусственный” в пределах, представленных на Рис. 4. Неравномерность частотной характеристики звукового давления установки“ Рот искусственный” в диапазоне частот 200-4000 Гц не более 6 дБ.

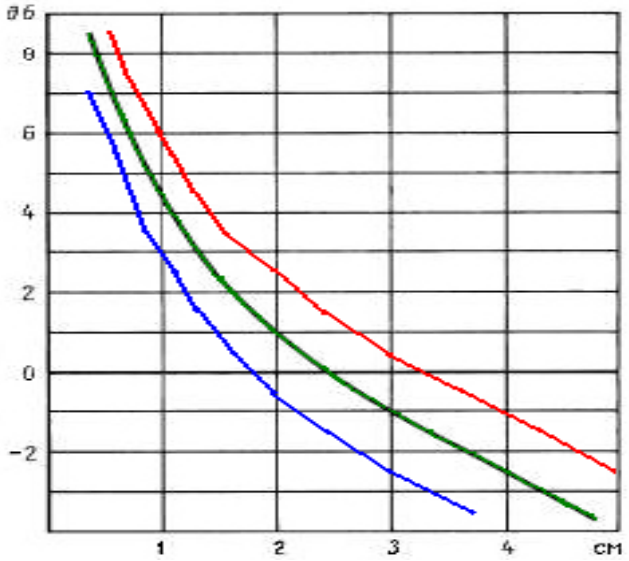


Рис. 4. Пределы изменения звукового давления развиваемого динамиком установки “Рот искусственный” в речевом диапазоне в зависимости от расстояния.

Результаты распознавания регистрируются и обрабатываются по каждому диктору или движку синтезатора. Отдельно обрабатываются данные по результатам, полученным в каждом из режимов (диктование, тот или иной метод синтеза речи).

3. Заключение

Тестирование системы распознавания команд в режиме диктования диктора очевидно дает оценки, наиболее приближенные к реальным, но одновременно является наиболее трудозатратным.

Тестирование системы распознавания команд в режиме воспроизведения аудиофайлов требует наличия большой предварительно подготовленной базы аудиоданных. Достоинством этой технологии является возможность многократно повторного точного воспроизведения условий эксперимента.

Тестирование системы распознавания команд в режиме синтеза команд по предварительно подготовленным аудиозаписям слов и воспроизведения синтезированных речевых команд требует небольшой предварительно подготовленной базы аудиоданных. Достоинством этой технологии также является возможность многократно повторного точного воспроизведения условий эксперимента.

Современное состояние синтезаторов русской речи позволяет проводить ограниченное тестирование, однако тестирование системы распознавания команд с использованием синтезаторов речи является для авиационных приложений наиболее перспективной и наименее трудозатратной технологией.

Литература

1. ГОСТ 17187-81 (СТ СЭВ 1351-78) “Шумомеры. Общие технические требования и методы испытаний”.
2. ГОСТ 8.257-84 “Шумомеры. Методика проверки”.
3. ГОСТ 20296-81 Самолёты и вертолёты гражданской авиации. Допустимые уровни шума в салонах и кабинах экипажа и методы измерения шума.
4. ITU-T P.50 Telephone Transmission Quality. Objective Measuring Apparatus. Artificial Voices.
5. ITU-T P.51 Telephone Transmission Quality. Objective Measuring Apparatus. Artificial Mouth.
6. ITU-T P.57 Telephone Transmission Quality. Objective Measuring Apparatus. Artificial Ears.
7. ITU-T P.58 Head and torso simulator for telephonometry.
8. ITU-T P.59 Artificial conversational speech.
9. ED-18 Audio Systems Characteristics and Minimum Performance Specifications. Aircraft Microphones (exept carbon). Aircraft Headsets, Headsets and Loudspeakers. Aircraft Audio Selector Panels and Amplifiers.

М.В. Прищепа, В.Ю. Будков

Разработка интерактивной модели речевого взаимодействия мобильного информационного робота с посетителями торговых центров[1]

Санкт-Петербургский институт информатики и автоматизации РАН,
Санкт-Петербург, Россия,
vishar2k6@bk.ru , budkov@iias.spb.su

1. Введение

Анализ последних научных публикаций показывает, что социальные роботы могут использоваться в качестве проводников в музеях, школьных учителей, для психологической помощи людям с ограниченными физическими возможностями, а также для ухода за детьми [1, 2]. В данной работе проводится разработка робота, способного предоставлять пользователям необходимую информацию в сложной и постоянно изменяющейся окружающей обстановке, например в торговых центрах. По сравнению со стандартными баннерами и информерами (картами) робот обладает большей информативностью и может предоставлять более качественные услуги. Робот обладает рядом преимуществ, таких как, например, гуманоидная или антропоморфная форма, позволяющая ему объяснить посетителю маршрут не только с помощью речи или схемы, но и с помощью жестов, также неоспоримым преимуществом является мобильность и автономность, что позволяет роботу проводить посетителя до точки назначения.

Социальные роботы должны работать в тесном контакте с людьми, и уметь общаться как вербальными, так и невербальными способами. Робот должен не только понимать речь и жесты человека, но и самостоятельно воспроизводить их. Идея создания социально-ориентированных роботов заключается в создании систем, с которыми человек мог бы взаимодействовать без предварительного обучения. Разработка таких систем ведется во многих странах, и получено уже достаточное количество результатов, которые говорят об успешности идеи в целом. В качестве примеров социальных роботов оказывающих информационные услуги можно отметить проекты Sage и Mel.

В проекте Mel [3] был создан робот-пингвин, способный распознавать предметы и людей, понимать и воспроизводить речь, а также активно пользоваться жестикуляцией. По запросу пользователя он может указать ему на местонахождение искомого объекта. За счет объединения различных способов коммуникации: жестов, направления взгляда, речи – обеспечивается эффективный двухсторонний диалог между человеком и роботом.

В проекте Sage [4] разрабатывался мобильный подвижный робот, оснащенный экраном и стереодинамиками, взаимодействующий с пользователями посредством речи и эмоций. Робот является полностью автономным и в настоящее время установлен в одном из залов музея естественной истории имени Карнеги, где предоставляет пользователям необходимую информацию и выступает в роли гида.

Способность вести естественный диалог, перемещаться вместе с посетителем и учитывать его предпочтения является основным преимуществом мобильного информационного робота. Наиболее актуальным местом применения подобных роботов считаются торговые центры, так как покупателям с каждым годом становится все сложнее ориентироваться в новых все расширяющихся комплексах, а непривычная новизна робота привлекает посетителей, и поэтому его применение в рекламных целях весьма эффективно [5].

[1] Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (ГК № НК-588П_4).

Для пользователей информационный робот является одним из представителей торгового центра, поэтому он должен быть дружелюбным, чтобы посетители чувствовали себя комфортно. Так как люди регулярно посещают торговые центры, то роботу необходимо запоминать постоянных посетителей и их предпочтения. Кроме того, робот должен вести себя естественно, и обладать чувством юмора для привлечения посетителей, а также быть в курсе последних важных новостей [5, 6].

2. Алгоритмическая модель поведения информационного робота

В состав разработанного мобильного информационный робота входит подвижная платформа и установленная на ней информационная стойка. Многомодальный интерфейс, разработанный ранее для стационарного информационно-справочного киоска, был использован при проектировании робота [7]. Для определения положения пользователя и речевых запросов применяется набор веб-камер, массив микрофонов и технологии дистанционного распознавания речи и слежения за лицом человека [8].Кроме стандартного графического интерфейса, реализованного посредством сенсорного экрана, для генерации речевых сообщений пользователю и слежения за его перемещением в ходе диалога применяется аудиовизуальный синтез речи «говорящая голова» [9].

В ходе исследования в первую очередь была предложена алгоритмическая модель информационного робота, описывающая основные режимы его работы. На рисунке 1 приведена схема модели поведения информационного робота.

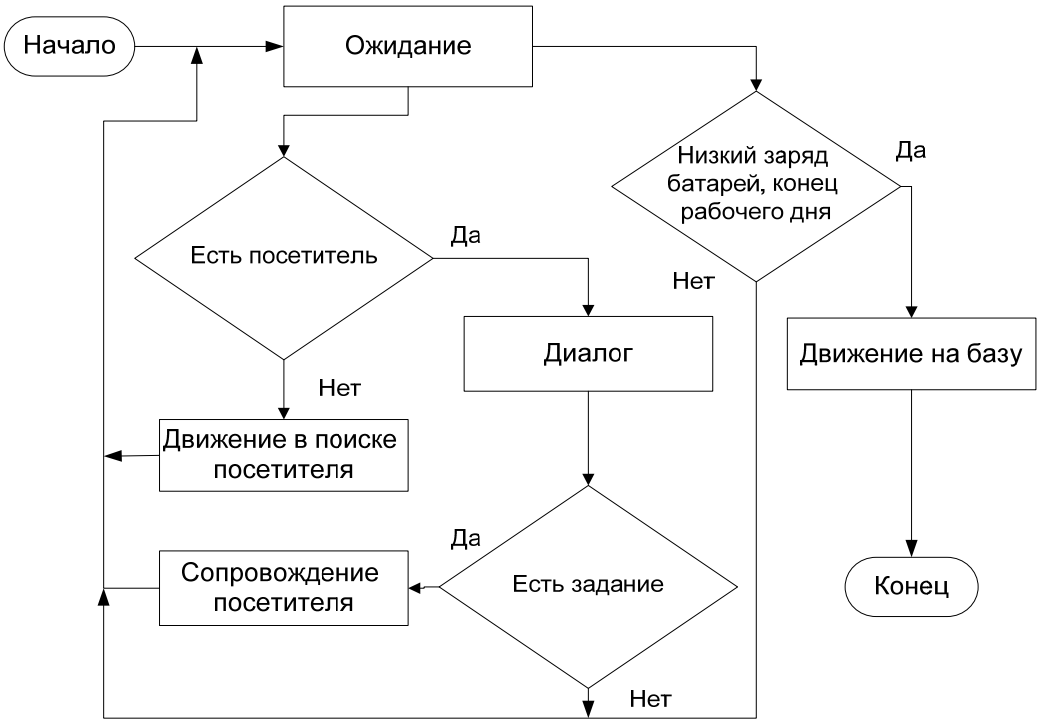


Рис.1. Алгоритмическая модель поведения информационного робота.

В этой модели поведения можно выделить несколько режимов работы робота – ожидание, поиск посетителя, диалог, сопровождение посетителя, движение на базу. В режиме ожидания аватар, в виде говорящей головы, следит за перемещением людей. Визуализация изменения направления взгляда головы создает впечатление у прохожего, что за ним следят, и привлекает его внимание к роботу, как только клиент оказывается в зоне речевого взаимодействия и остается в ней определенное время, робот озвучивает приветствие и начинает диалог. Также возможен другой стиль поведения в режиме ожидания - робот прокручивает рекламные ролики на мониторах, и в тоже время следит за окружающей обстановкой в поиске пользователей. Если в течение определенного времени

пользователь не появился, то робот начинает движение по заданному маршруту в поисках посетителя. Если в ходе движения пользователь не появился, то робот снова переходит в режим ожидания. Когда появляется посетитель, робот переходит в режим диалога и предоставляет необходимую информацию пользователю. Пользователь может попросить робота проводить его к нужному месту, в этом случае робот переходит в режим сопровождения посетителя, и после того выполнит задание снова переходит в режим ожидания. Если в ходе работы заряд батареи становится низким, то в этом случае робот переходит в режим движения на базу; то же самое происходит в конце рабочего дня.

Наиболее важным режимом работы является режим диалога. В режиме диалога производится захват кадра и его анализ с помощью алгоритма Хаара, обеспечивающего поиск лица. Если на изображении обнаруживается несколько лиц, то выбирается ближайшее. Затем производится расчет расстояния до него, учитывая стандартные пропорции лица. Чтобы объект оказался в зоне речевого взаимодействия, робот перемещается в такую позицию относительно человека, чтобы расстояние между ними составляло не более 2 метров, а отклонение от центра камер не превышает 20 градусов. В ходе диалога робот продолжает анализировать положение человека и старается перемещаться таким образом, чтобы удерживать человека в зоне речевого взаимодействия. Также при запросе пользователя робот может указать необходимое направление, или самостоятельно проводить посетителя в нужное место. На рисунке 2 приведена схема управления роботом в режиме диалога.

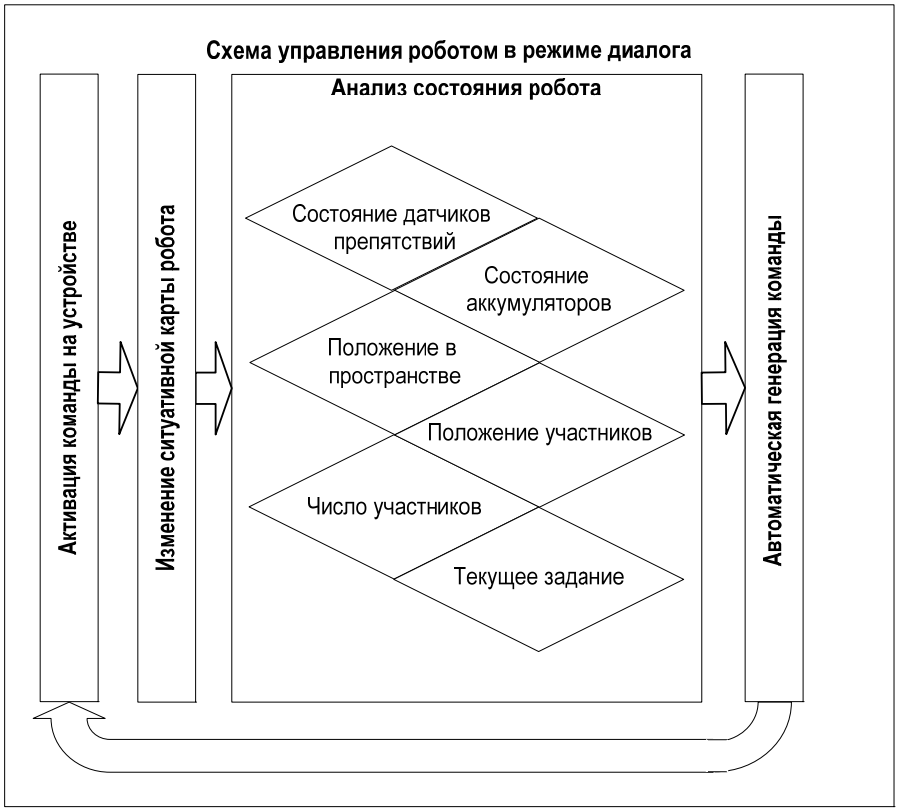


Рис. 2. Схема управления информационным роботом во время диалога.

При поступлении команды фиксируются сведения о наличие в этот момент пользователей и их положение. Когда количество пользователей превышает одного человека, приоритетным пользователем автоматически признается тот, кто находится ближе всего к роботу. При поступлении задания от пользователя проверяется наличие других заданий (например, движение на базу), если задание уже существует, то робот извиняется и просит подождать, пока не выполнит текущее задание, которое является приоритетным, так как

было отдано раньше по времени. Также опрашиваются датчики препятствий и состояние аккумуляторов. Если нет никаких ограничений, то автоматически генерируется команда для исполнительных устройств. При срабатывании датчиков вырабатывается команда остановки электроприводов, чтобы избежать столкновения с препятствием. В случае окончания заряда аккумуляторов, робот оповещает об этом пользователей (если они в данный момент находятся около него) и начинает движение на базу для подзарядки. Схема обработки команд пользователя, представленная на рисунке 2, допускает автоматические действия робота в соответствии с текущей ситуацией [10].

Рассмотрим разработанную диалоговую модель взаимодействия информационного робота с посетителем. Необходимо отметить, что модель взаимодействия должна отвечать таким условиям как: (1) информативность – предоставленная информация должна иметь актуальность; (2) естественность – диалог системы с пользователем должен происходить на естественном для пользователя языке и не требовать от него каких-либо специальных знаний и навыков; (3) лаконичность – количество предоставляемой информации должно быть строго таким, какое нужно пользователю в данный момент, и не должно содержать лишние или ненужные данные. Разработанная модель взаимодействия основывается на использовании речи в качестве главного способа коммуникации, а также использовании жестов в качестве вспомогательного. Модель предназначена для взаимодействия пользователей с информационной стойкой в таких сложных условиях как торговые центры. Стойка, находясь в помещении торгового центра, предоставляет информационные услуги по нахождению нужных магазинов, категорий товаров, или служебных помещений. На рисунке 3 представлена схема диалоговой модели взаимодействия информационного робота с посетителями.

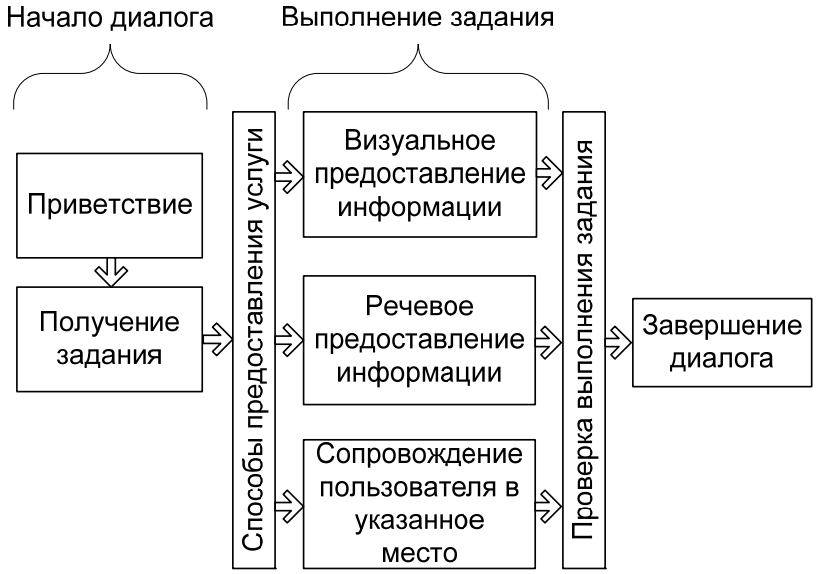


Рис.3. Схема диалоговой модели информационного робота.

Как видно на схеме модель представляет собой 3 основных блока: начало диалога, выполнение задания и завершение диалога. На первом этапе происходит знакомство пользователя с системой. Информационный робот представляется и предлагает перечень услуг, которые он может оказать посетителю: указание местоположения магазинов (по названиям), указание магазинов по категориям товаров (например, обувь - названия всех обувных магазинов в данном торговом центре), новости и информация о скидках и акциях проводимых в магазинах торгового центра, возможность проводить пользователя в нужную ему точку торгового центра. Далее пользователю предлагается выбрать категорию услуг и дать задание информационному роботу. После получения задания робот приступает к их выполнению. По запросу пользователя он может озвучить запрашиваемую информацию, с

помощью аудиовизуального синтеза речи [9]. Если посетитель предпочитает визуальное предоставление информации то информационный робот выведет ее на экран в виде фотографии магазина, названия и каталога его продукции. Аудиовизуальная модель человека может указать требуемое направление жестами [11]. Также робот способен проводить посетителя до места назначения по его просьбе. После того как задание выполнено робот прощается с посетителем и переходит в ждущий режим.

3. Заключение

Разработка диалоговых систем является комплексной задачей и при разработке структуры диалога, прежде всего, учитываются знания экспертов предметной области. При этом не только структура диалога, но и сами фразы, задаваемые диалоговой системой, должны быть тщательно подобраны, чтобы сократить время диалога до минимума [12, 13]. Так как цель многих диалоговых систем обеспечить пользователя интересующей информацией, то возможно использовать в качестве критерия оптимальности стратегии диалога, среднее время успешного взаимодействия, т.е. когда пользователь проходит все этапы диалога, и по завершению получает необходимую информацию.

Литература

1. Nass, C., Steuer, J. et al. Computers are social actors. In. Proc. of the CHI'94 Conference on Human Factors in Computing Systems, Publ by ACM, NY, USA, 1994, pp. 72-78.
2. Iossifidis, I., C. Theis, et al. Anthropomorphism as a pervasive design concept for a robotic assistant. In Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003, Vol.3, pp. 3465-3472.
3. Sidner, C. L., C. Lee. Robots as laboratory hosts // Interactions 12 (2), 2005, pp. 24-26.
4. Nourbakhsh, I. R., J. Bobenage, et al. Affective mobile robot educator with a full-time job // Artificial Intelligence 114 (1-2), 1999, pp. 95-124.
5. Kanda, T. Shiomi, M., Miyashita, Z., Ishiguro, H., Hagita, N. An Affective Guide Robot in a Shopping Mall. In Proc. of Conference on Human-Robot Interaction 2009, 2009, pp. 173-180.
6. Bartneck, C., Okada, M. Robotic User Interface. In Proc. of the Human and Computer Conference (HC-2001), Aizu, 2001, pp. 130-140.
7. Ронжин А.Л., Карпов А.А., Кагиров И.А. Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информационно-управляющие системы, Вып. 42, т. 5. — СПб.: ГУАП, 2009, С. 32-38.
8. Karpov A.A., Ronzhin A.L., Information Enquiry Kiosk with Multimodal User Interface // Pattern Recognition and Image Analysis, Moscow: MAIK Nauka/Interperiodica, Vol. 19, № 3, 2009, pp.546-558.
9. Карпов А., Ронжин А., Лобанов Б., Цирульник Л., Железны М. Разработка бимодальной системы аудиовизуального распознавания русской речи // Информационно-измерительные и управляющие системы, Москва, № 10, Т. 6, 2008, С. 58-62.
10. Будков В.Ю., Прищепа М.В., Ронжин А.Л. Анализ способов взаимодействия с мобильным информационным роботом. Завалишинские чтения: Сборник докладов / ГУАП, СПб, 2009, С.12-19.
11. Hruz M., Campr P., Karpov A., Santemiz P., Aran O., Zelezny M. Input and Output Modalities Used in a Sign-Language-Enabled Information Kiosk. In Proc. of the 13-th International Conference SPECOM'2009, St. Petersburg, 2009, pp. 113-116.
12. Bersen, N. O., Dybkjaer, H. & Dybkjaer, L. Designing interactive speech systems: From first ideas to user testing. Springer-Verlag, 1998.
13. Ронжин А.Л., Карпов А.А., Ли И.В. Речевой и многомодальный интерфейсы. - М.: Наука, 2006 - (Информатика: неограниченные возможности и возможные ограничения), 173 с.

С.В. Глазков

Разработка интерактивных приложений с многомодальным интерфейсом для гетерогенных мобильных устройств[1]

Санкт-Петербургский институт информатики и автоматизации РАН,
Санкт-Петербург, Россия,
s.glazkov@gmail.com

1. Введение

Многомодальные интерфейсы обеспечивают взаимодействие на основе различных коммуникационных каналов, обрабатывая два и более пользовательских способа ввода информации совместно с мультимедийной системой вывода информации. Модальности – это те каналы, которые люди могут использовать для взаимодействия с людьми или техникой, например визуальный, звуковой, тактильный каналы, вкус, запах и проприоцепция (ориентация тела) [1]. В отличие от традиционных интерфейсов на основе клавиатуры и мыши или одномодальных интерфейсов, многомодальные системы обеспечивают более гибкое использование входных потоков информации. Это дает возможность человеку выбирать наиболее удобный способ передачи/приема информации.

В таблице 1 представлены основные типы модальностей, используемые человеком при коммуникации с людьми, на базе которых строятся варианты многомодальных интерфейсов [2]. Визуальная и звуковая модальности при построение интерфейсов применяются более активно. Дополнительно выделяют чувства присутствия и соприсутствия, связанные с поведением людей, находящихся непосредственно в коллективе или участвующих в мероприятиях удаленно, используя телекоммуникационные средства связи [3]. В приложениях в области медицины и безопасности также анализируются другие биометрические показатели: отпечаток пальца, форма ладони, термограмма лица, рисунок сосудов глазного дна и другие.

Таблица 1. Основные типы модальностей и примеры технологий их обработки

Органы чувств	Типы модальностей	Примеры модальностей	Примеры технологий и средств обработки
Глаза	Визуальная	Жесты рук, тела, движения губ, направление взгляда, анимационный объект	Анализ изображений, технологии виртуальной реальности
Уши	Звуковая	Речь, звуки, мелодии	Распознавание и синтез разговорной речи
Кожный покров	Тактильная	Прикосновение, вибрация	Сенсорные панели и тактильные дисплеи
Нос	Обонятельная	Запах	Методы химического анализа и синтеза
Язык	Вкусовая	Вкус	
Вестибулярный аппарат	Вестибулярная	Положение тела	Системы на базе акселерометров и гироскопов
Суставы, нервная система	Проприоцептивная	Относительное положение частей тела и их движения	Системы с экзоскелетом

[1] Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (ГК № П2360) и проекта РФФИ (№10-08-00199-а.

В зависимости от использованных входных и выходных модальностей выделяют несколько основных типов многомодальных интерфейсов (речь+жесты, речь+чтение по губам, направление взгляда+указание+речь и т.д.). В данной работе анализируются особенности программно-аппаратного обеспечения мобильных устройств, которые главным образом влияют на построение многомодальных и речевых интерфейсов.

2. Особенности программно-аппаратного обеспечения мобильных устройств

При разработке приложений для мобильных устройств необходимо учитывать его ограничения по размерам, весу, времени автономной работы, а также производительности и объему встроенной памяти. Небольшой размер таких устройств, а также условия их использования ограничивают возможности применения привычного интерфейса (клавиатура и мышь). Разместить удобную полноразмерную клавиатуру в мобильном устройстве достаточно сложно, а сделать ее пригодной для использования методом слепой печати вообще не предполагается в ситуациях, когда нет возможности расположить мобильное устройство на столе или других твердых поверхностях, например, за рулем автомобиля или в общественном транспорте в часы пик. Таким образом, наиболее предпочтительными входными модальностями для взаимодействия с мобильными устройствами являются аудиовизуальные (как наиболее приближенные к межчеловеческому общению).

Производительность мобильных устройств несоизмерима с производительностью персональных компьютеров, а тем более с производительностью специализированных серверов. Для решения сложных комплексных задач, наиболее эффективно применять сетевую (клиент-серверную) модель обработки данных. Реализация сложного многомодального интерфейса, в котором в качестве основных модальностей выступают звук и изображение, с возможностью обработки параллельно более одной модальности не представляется возможным только средствами мобильного устройства, но возможна в распределенной системе, где мобильное устройство выступает в первую очередь сборщиком данных, а вся основная обработка выполняется на сервере. В этом случае применение сложных алгоритмов распознавания речи и видеопотока, такие как скрытые Марковские модели, нейронные сети, алгоритмы глобальной реконструкции возможно без каких либо ограничений. Но при использовании такого подхода необходимо учитывать следующие особенности мобильных устройств, а также сетей связи и беспроводной передачи данных:

- *Тактовые частоты процессоров.* Производительность процессора мобильного устройства может быть крайне мала. Это замедляет первичную обработку информации и скорость отправки и приема данных, а также скорость шифрования/дешифрования данных, пересылаемых по защищенным коммуникационным каналам связи.
- *Полу- и полнодуплексная радиосвязь.* Некоторые устройства не поддерживают полнодуплексный режим, т.е. не могут одновременно передавать и принимать данные. Это увеличивает задержки и отражается на скорости реагирования устройства на действия пользователя.
- *Запаздывание.* Большинство мобильных устройств использует энергосберегающие протоколы. Кроме того, существуют законодательные ограничения на удельную степень поглощения (Specific Absorption Rate, SAR). SAR - это мера биологического воздействия радиочастотного излучения, характеризующая безопасность сотовых телефонов и других мобильных устройств. Эти ограничения дополнительно увеличивают время запаздывания мобильных устройств.
- *Пропускная способность.* Хотя мобильные устройства GPRS (General Packet Radio Services) могут работать на теоретически максимальных скоростях 171,2 килобит в секунду (Кбит/с), многие телефоны все еще используют технологии

Circuit Switched Data, поддерживающие скорость передачи данных не более 9,6 Кбит/с. Это значит, что мобильные устройства особенно чувствительны к издержкам, связанным с работой протоколов, и к объему передаваемых данных. В данном исследовании первоначально планируется использовать сеть беспроводной связи по технологии WiFi с максимальной теоретической скоростью передачи до 54Мбит/с, но, тем не менее, ограничения должны быть учтены на этапе проектирования.

- *Негарантированная возможность соединений.* Мобильное устройство может перемещаться из зоны устойчивого охвата в зону частичного охвата или оказываться там, где вообще связь отсутствует. На качество связи влияют высокие здания, мосты и другие объекты. Поэтому необходимо следить за наличием соединения и при его отсутствии переключать интерфейс устройства в локальный режим работы, уведомляя об этом пользователя.

Также необходимо уделить особое внимание таким аспектам как: минимизация времени отклика интерфейса и использование различных методов шифрования для защиты пользовательских данных.

3. Особенности взаимодействия пользователя с мобильным устройством

В целом можно сказать, что трудности освоения мобильных компьютерных технологий обусловлены тем, что с их помощью пользователи решают все более сложные задачи, в то же время при разработке интерфейсов этих систем психологические знания применяются недостаточно.

Необходимо отметить, что в современных средствах разработки приложений практически не автоматизирован процесс комплексного тестирования и оценки параметров качества пользовательского интерфейса (с использованием формальных и эвристических методов). Среди существующих методов доминируют неформальные подходы, при этом значителен субъективизм оценок, общей является попытка выделить (с помощью обширного тестирования пользователями и самих пользователей) общие, типовые характеристики присущие качественным интерфейсам и на их основе создать методику оценки эффективности графического пользовательского интерфейса. Как правило, существующие методики рассматривают какой-то один аспект оценки эффективности графического пользовательского интерфейса, выделяя при этом единственный наиболее значимый критерий.

В основе многих теоретических, практических и эмпирических исследований, посвященных разработке моделей коммуникативного взаимодействия человека с компьютером, лежит классический вопрос – что разработчику необходимо знать о пользователе, чтобы организовать эффективное человеко-машинное взаимодействие. Взаимодействие пользователей с мобильной компьютерной системой может быть облегчено, если форма и структура программного обеспечения будут согласовываться со свойственными человеку особенностями восприятия, мышления и поведения.

Все вышесказанное относится к пользовательским интерфейсам в целом без акцента на мобильные устройства и их особенности. Далее будут рассмотрены проблемы связанные исключительно с мобильными устройствами.

Пользователи сталкиваются с растущим разнообразием мобильных устройств, их операционными системами и, соответственно, пользовательскими интерфейсами. Удобство использования того или иного мобильного устройства определяется, чаще всего, привычками пользователя. Интерфейсы с развитием аппаратной составляющей мобильных устройств усложняются, появляется большое количество настроек, приложений, периферийных устройств. Дисплеи устройств невелики, и не позволяют разместить все часто используемые настройки на одном экране, в интерфейсе появляются дополнительные рабочие столы, многочисленные переходы по различным меню, все это не делает интерфейсы проще. Зачастую, владельцы мобильных устройств даже не знают

обо всех возможностях предоставленных им мобильным устройством или не желают пользоваться ввиду неудовлетворительной или неудобной реализации некоторых функций.

Эргономика мобильных телефонов, главным образом направлена на обеспечение голосового взаимодействия пары пользователей. Цифровая клавиатура мобильного телефона изначально была предназначена для набора номера абонента, но с появлением дополнительных сервисов предоставляемых операторами связи, позволила набирать и текстовые сообщения. Буквы на таких клавиатурах расположены в алфавитном порядке. Путем повторного нажатия одной и той же клавиши цифровой клавиатуры можно выбрать нужный символ, назначенный на эту клавишу. Скорость ввода текстовой информации при использовании вышеописанного метода составляет в среднем 8 слов в минуту. Более опытные пользователи мобильных устройств способны набирать в среднем до 20 слов в минуту благодаря технологии предикативного ввода T9 [4], но при таком вводе полностью задействуется визуальное внимание человека. С появлением сенсорных экранов, кнопочные клавиатуры стали заменять программными аналогами, стали появляться полноразмерные QWERTY клавиатуры, отображающиеся на дисплее. Ввод хоть и стал более привычным, в сравнении с настольным компьютером, но остался очень медленным.

Было проведено большое количество работ по ускорению ввода данных в мобильных устройствах [5, 6, 7], большинство из них направлено на адаптацию существующих способов ввода от стационарных компьютеров к мобильным устройствам [8]. К сожалению, такой подход пока не дал каких-либо существенных результатов. Скорость ввода данных в мобильных устройствах оставляет желать лучшего, существующие способы ввода требуют от пользователя практически полного внимания и невозможно набирать текст, находясь в движении или за рулем автомобиля, следовательно, существующие интерфейсы накладывают ограничение и на мобильность пользователя. Чтобы комфортно написать текстовое сообщение или электронное письмо придется, по крайней мере, остановиться. Все это лишь подтверждает мнение, что для мобильных устройств, необходимы принципиально иные интерфейсы и способы ввода. Эти интерфейсы должны обеспечить взаимодействие пользователя с мобильным устройством, сняв следующие ограничения: (1) невозможность ввода данных в движении; (2) невозможность ввода данных без необходимости смотреть на экран устройства; (3) невозможность использования устройства без прямого контакта (т.е. если телефон находится в кармане, воспользоваться его функциями не удастся до тех пор, пока он не будет извлечен из кармана). В следующем разделе рассмотрим основные проблемы, связанные с внедрением речевого интерфейса в мобильные устройства.

4. Анализ вариантов речевого интерфейса для мобильных устройств

Основной входной модальностью в мобильном устройстве является речь. По прогнозам крупной аналитической компании, мировой рынок систем распознавания речи (ASR) для мобильных телефонов в следующие 5 лет вырастет в 3 раза [9]. В то же время, самой крупной областью применения речевых технологий по-прежнему будет оставаться сегмент автомобильной телематики. Ожидается, что объем мирового рынка систем распознавания речи вырастет с $32,7 млн. в 2009 г. до $99,6 млн. в 2014 г. Примерно теми же темпами будет расти и рынок ASR для автомобильных телематических систем: с $64,3 млн. в 2009 г. до $208,2 млн. в 2014 г. «Рост популярности голосового интерфейса в телефонах будет расти по мере того, как все большее число их владельцев сталкиваются с необходимостью использовать мобильник в ситуациях, когда руки и глаза заняты», - говорят специалисты [9].

Несмотря на кажущуюся простоту, по сравнению с системами распознавания речи (системами документирования устных выступлений, например) для стационарных компьютеров, приложения, управляющие мобильными устройствами с помощью речи,

могут использоваться ежедневно в любом месте с различными типами фоновых шумов, поэтому основной проблемой распознавания речи в мобильных устройствах является не размер словаря, хотя и он имеет большое значение, а помехоустойчивость системы распознавания [10]. С одной стороны должны учитываться как стационарные помехи, такие как разговоры, музыка, шум движения автомобиля или поезда, а с другой случайные звуки, такие как стук захлопнувшейся двери, звонок мобильного, или смех, которые могут быть распознаны как команда. Также распознавание должно корректно работать в независимости от типа микрофона, на любом расстоянии и под любым углом между ртом и микрофоном, а также при использовании handset и handsfree устройств. И все это без необходимости переобучать систему распознавания. Поскольку потребность в речевом интерфейсе обусловлена необходимостью использовать мобильное устройство без участия рук и глаз, дополнительный диалоговый речевой интерфейс, с повторением при необходимости введенной команды, позволит помочь избежать потери внимания, и пользователю не потребуется брать устройство в руки.

Выбор подходящего типа системы распознавания для мобильного устройства особенно важно. Далее рассмотрим базовые варианты систем, их преимущества и недостатки [11]:

Пословное распознавание против распознавания слитной речи. Пословные распознаватели позволяют выполнять распознавание одного слова за один цикл распознавания. Чтобы распознать несколько слов таким распознавателем, пользователь должен делать паузу перед каждым следующим словом, чтобы заканчивать один цикл распознавания и начинать новый, такая речь неестественна и потому неудобна для пользователя. Распознаватели слитной речи позволяют пользователю произносить последовательность слов без искусственных пауз между словами. Основная проблема такого способа распознавания – объединение слов, в тех случаях, когда их произношение отличается от написания, как, например, последовательность цифр. Границы слов часто исчезают и слова сливаются при распознавании. Тем не менее, поскольку слитная речь – естественная форма межчеловеческого общения, то распознавание слитной речи является наиболее предпочтительным в мобильных устройствах.

Дикторозависимое или дикторонезависимое распознавание. Часто пользователи мобильных устройств, прежде чем пользоваться встроенной системой распознавания команд должны ее натренировать, например, чтобы воспользоваться голосовым набором телефона из записной книжки мобильного телефона, необходимо на вызываемый номер установить голосовую метку. Однако, несмотря на то, что тренировка требует от пользователя дополнительных действий, она имеет одно неоспоримо преимущество – независимость от языка распознавания. То есть, любой пользователь может натренировать свой телефон на любой желаемый язык. Этот факт значительно упрощает внедрение голосового управления в различные типы устройств по всему миру. Дикторонезависимые системы распознавания обучаются на большом количестве образцов записей дикторов. Пользователь может сразу же начать пользоваться такой системой. Недостатком такого подхода является невозможность охватить все возможные варианты произношения, поэтому некоторые команды могут распознаваться с меньшей точностью. В таком случае, необходимо предоставить пользователю возможность адаптации системы, чтобы улучшить распознавание исключительно тех команд, которых пожелает пользователь.

Малый или большой словарь распознавания. Для реализации системы управления мобильным устройством с помощью речи небольшого словаря до 100 слов будет вполне достаточно. Обычно в таких системах используют пословные распознаватели, то есть один словарный элемент – одно слово. Требуемый объем памяти такого рода систем ограничен, а вычислительной мощности мобильного устройства будет вполне достаточно для их реализации, что является бесспорным преимуществом перед системами с большими словарями распознавания. Системы с большим словарем распознавания, позволяющие диктовать текстовые короткие сообщения или электронные письма, могут

содержать до 100 000 слов. Распознавание в таких системах построено на выделении из речевого сообщения фонем - базовых единиц, из которых состоит речь на естественном языке. Затем распознанные фонемы используются для построения слов на основе словаря. Сложность подобных систем пока не позволяет реализовывать подобные системы на мобильных устройствах, поскольку их вычислительных мощностей недостаточно для решения столь комплексных задач, а требуемые объемы памяти могут исчисляться гигабайтами.

5. Заключение

Многомодальные интерфейсы в целом и речевые в частности – несомненно, будущее мобильных устройств. С развитием аппаратной составляющей мобильных устройств, станет возможным реализовывать системы, функционирование которых в настоящее время возможно только на стационарных компьютерах или специализированных серверах. Сейчас существует возможность реализовывать диалоговые системы, применяя адаптированные под мобильные телефоны алгоритмы с небольшим словарем и пословным распознаванием, а с современными темпами развития информационных технологий и аппаратной базы станет возможно применение более сложных моделей обработки речи и других естественных модальностей.

Литература

1. Oviatt, S. L. Mutual disambiguation of recognition errors in a multimodal architecture. Proceedings of the Conference on Human Factors in Computing Systems (CHI’99), New York: ACM Press, 1999. – pp. 576-583.
2. Шиффман Х. Ощущение и восприятие. СПб:Питер, 2003. – 928 с.
3. Op den Akker, R., Hofs, D., Hondorp, H., Akker, H., Zwiers J., Nijholt, A. Supporting Engagement and Floor Control in Hybrid Meetings. Springer, LNAI 5641, 2009. – pp. 276-290.
4. James, C.L., Reischel, K.M. Text input for mobile devices: comparing model prediction to actual performance. In Proc. of the SIGCHI conference on Human factors in computing systems. Seattle, WA: ACM Press, 2001, pp. 365-371.
5. Harrison, C., Tan, D. Morris, D. Skinput: Appropriating the Body as an Input Surface. In Proc. of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, April 10 - 15, 2010. CHI '10. ACM, New York, NY, 2010 vol.1 pp.453-462.
6. Harrison, C., Hudson, S. Minput: Enabling Interaction on Small Mobile Devices with High-Precision, Low-Cost, Multipoint Optical Tracking. In Proceedings of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, April 10 - 15, 2010). CHI '10. ACM, New York, NY, 2010, vol.1 pp.1661-1664.
7. Schwarz, J., Harrison, C., Mankoff, J., Hudson, S. E. 2010. Cord Input: An Intuitive, High-Accuracy, Multi-Degree-of-Freedom Input Method for Mobile Devices. In Proc. of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, April 10 - 15, 2010, CHI '10. ACM, New York, NY vol.1 pp.1657-1660.
8. Sears, A., Zha, Y. Data Entry for Mobile Devices Using Soft Keyboards: Understanding the Effects of Keyboard Size and User Tasks // International Journal of Human-Computer Interaction, Vol. 16, Issue 2, 2003, pp. 163 – 184.
9. http://www.cellular-news.com/story/37565.php
10. Dobler, S. Speech recognition technology for mobile phones // Ericsson Review no. 03, 2000.
11. Benesty, J., Sondhi, M., Huang, Y. (eds.) Springer Handbook of Speech Processing. Springer, 2008, 1176 p.

М.А. Мясоедова, Д.Н. Душкин, З.П. Мясоедова, Н.В. Петухова, М.П. Фархадов

Разработка интернет портала «Сурдосервер» с ресурсами русского жестового языка

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия
serena@ipu.ru; mais@ipu.ru

1. Введение

Помощь людям с ограниченными возможностями, вовлечение их в трудовую и общественную жизнь является важной социальной задачей любой страны, любого общества. В Институте проблем управления Российской Академии наук при поддержке Российского Фонда фундаментальных исследований был выполнен проект по созданию интернет портала «Сурдосервер», предназначенный для помощи людям с нарушениями слуха.

2. Виды жестовой речи

В ходе выполнения проекта был произведен анализ русского жестового языка и выявлены его особенности. Существуют две разновидности жестовой речи: разговорная и калькирующая. Каждая из этих разновидностей имеет свое функциональное назначение и свою лингвистическую структуру.

Разговорная жестовая речь применяется в условиях непринужденного межличностного общения. Разговорный жестовый язык является родным языком врожденно глухих или рано оглохших людей, выросших в семье глухих родителей либо достаточно рано попавших в коллектив глухих. Он также служит инструментом повседневной коммуникации неслышащих людей и смешанных групп, в которые входят как люди с нарушением слуха, так и их слышащие друзья и родственники. Разговорная жестовая речь не является калькой вербальной разговорной речи, но оба эти типа разговорной речи имеют ряд общих черт: ограниченность словаря, игнорирование грамматических правил, сжатость, конситуативность, использование слов с «отсылочным» значением, типа эта вещь, эта штука, замещающих любую лексему [1-3].

3. Особенности разговорного жестового языка

Особенности разговорного жестового языка обусловлены его ориентацией на повседневное общение. Вот основные из них:

1. Ограниченность словаря. Поскольку этот язык используется для ведения разговоров о повседневной жизни, в нем не используются некоторые довольно распространенные жесты, такие, например, как жесты для обозначения головы, носа, руки и т.п. Вместо них используются указательные жесты. Вообще для разговорной жестовой речи очень характерно использование указательных жестов. Отсутствуют также жесты, обозначающие специальные термины, не характерные для повседневного общения. Они при необходимости передаются калькирующей речью или дактильной азбукой.

2. Морфология разговорного жестового языка. Жестовая речь не выражает склонений, спряжений, времени и других форм. Выражение этих форм может реализовываться по-разному, однако в основе этих реализаций лежит принцип «от значения к форме». Например, множественное число существительных можно выразить как последовательность жеста «существительное», например «книга», и жеста «много», либо как повторение несколько раз жеста для обозначения существительного. Аналогично выражается время глагола: либо как последовательность жеста, обозначающего глагол, и жеста «вчера», «давно» для прошедшего времени или жеста «завтра» для будущего времени, либо как основной жеста для выражения глагола и присоединения к нему жестов «было» или «будет».

3. Инфлексия (изменение способа исполнения жеста). Инфлексия очень распространена в разговорном жестовом языке. С помощью инфлексии может выражаться вид глагола (совершенный, несовершенный), лицо, число, законченность действия, субъектно-объектные отношения и многое другое. Например, жест в значении «покупать» исполняется медленно и повторяется, а тот же жест в значении «купить» – быстро и однократно. Движение от себя трактуется как действие от первого лица (объекта), а движение к себе – как действие второго лица (субъекта), т.е. если, например, жест «смотреть» исполнен от себя, то он означает «смотрю», а тот же жест, исполненный к себе, означает «смотришь на меня».

4. Расчлененность. Важной особенностью разговорной жестовой речи является значительно более частое использование расчлененности при передаче смысла лексем. Скажем, лексема «азербайджанец» строится как сочетание жестов «Азербайджан» и «человек», и такая структура чрезвычайно характерна для жестовой речи вообще, в то время, как в вербальной разговорной речи примеры использования расчлененности, например «что надеть» вместо «одежда», значительно более редки.

5. Конструкция предложения. Конструкция предложения в жестовой речи имеет свою специфику: обычно на первое место ставится предмет, затем его качество, после объекта указывается действие, которое на него направлено. Мимика часто выступает как семантическая единица высказывания.

6. Многозначность жестов. Для разговорной жестовой речи характерно использование одного жеста для выражения деятеля, действия и результата действия, например одним и тем же жестом передаются слова «писать», «письмо» и «писатель», или слова «топор» и «рубить». Конкретное значение жеста определяется контекстом.

7. Связь смысла высказывания с ситуацией и контекстом. В жестовой разговорной речи очень распространено свободное соединение жестов в сложные конструкции, где связи между членами можно понять, лишь зная ситуацию.

Таким образом, формы словесной устной речи не могут быть перенесены на разговорный жестовый язык. Разговорная жестовая речь – это самобытный язык, позволяющий выразить любые смыслы и отношения между смыслами, но традиционные лингвистические модели русского языка к ней не могут быть применены. Морфология, фразеология, идиоматика русского жестового языка изучены недостаточно.

4. Калькирующая жестовая речь

Калькирующая жестовая речь является кинетической копией русской литературной речи. С помощью калькирующей речи переводятся телевизионные новости, выступления на собраниях, лекции. Калькирующая речь не имеет собственной грамматики – построение жестового высказывания отражает порядок построения вербального предложения. В состав лексических единиц калькирующей речи входят собственно жесты, слова, представляемые дактилем, и лексемы, включающие жест и несколько дактилем. Для перевода слов, которые имеют эквивалент в жестовом языке, используются жесты. Жестам соответствуют слова в основной форме (именительный падеж существительных, инфинитив глаголов). Перевод имен собственных, терминов и незнакомых слов осуществляется с помощью дактильной азбуки. Сочетание жеста с дактилем используется для добавления окончаний и предлогов. Необходимым условием владения калькирующей жестовой речью является хорошее развитие словесной речи у людей с нарушениями слуха. Чтобы достичь хорошего владения словесной речью, необходимо начинать обучение ей детей с раннего возраста. Лишь наличие сформированных связей между значениями слов и соответствующими им жестами, знание грамматики вербального русского языка и лексики калькирующей речи и богатый словарный запас даст людям с нарушением слуха возможность понимать калькирующую речь. К сожалению, глухие от рождения люди часто не отвечают этим требованиям и плохо понимают высказывания, передаваемые калькирующей жестовой речью.

5. Национальные жестовые языки

В настоящее время в каждой стране имеется своя форма жестового языка, а иногда и несколько разных форм. Самым распространенным языком является американский жестовый язык ASL, им пользуются более полумиллиона человек. Следом по распространенности идут испанский и итальянский жестовые языки. Жестовый язык Великобритании отличается от американского жестового языка. Различны жестовые языки Китая и Тайваня. Два разных жестовых языка существуют в Австрии, в Ирландии и некоторых других странах. Многие языки имеют диалекты, вплоть до того, что каждая компактная группа глухих людей и даже каждая семья общаются на своей разновидности языка. Несмотря на национальные и диалектные отличия, знаковые языки для глухих все же имеют общие черты, и глухие люди из разных стран гораздо быстрее начинают понимать друг друга, чем слышащие люди. В 1975 году Всемирная федерация глухих для официальных церемоний приняла международный словарь для глухих, который включает в себя 1500 наиболее легко понимаемых знаков, взятых из различных национальных систем.

6. О проблемах перевода на жестовый язык

Машинный перевод на жестовый язык сталкивается с целым рядом проблем. Прежде всего это те же проблемы, которые имеют место при любом компьютерном переводе с одного языка на другой: выявление основного понятийного фонда языков, составление достаточно представительных словарей понятий, слов, фразеологических словосочетаний, концептуальный анализ текстов, адекватность перевода. В дополнение к этому существуют и специфические проблемы: пространственный характер жестового языка, значительные различия временных характеристик, недостаточная изученность структуры жестового языка [4]. Все это затрудняет разработку компьютерных переводчиков.

Были рассмотрены различные подходы к построению таких переводчиков для русского языка. По результатам анализа были сделаны следующие выводы:

- более простым в реализации представляется перевод с вербального русского языка на калькирующий жестовый язык. Недостатком этого варианта является ограниченность аудитории, которая сможет им пользоваться, поскольку глухие от рождения люди и люди, рано потерявшие слух, плохо владеют словесным языком. В то же время именно эти люди в наибольшей степени нуждаются в переводе текстов в понятную им форму;
- более целесообразным является перевод на разговорный жестовый язык, поскольку это будет удовлетворять нужды всего сообщества людей с нарушением слуха. Недостатком этого варианта является сложность реализации, недостаточная изученность разговорного жестового языка, многообразие диалектов.
- представляется целесообразным осуществлять поэтапную разработку сурдопереводчика, начав с составления единого словаря жестового языка, с обеспечения как можно более широкого доступа аудитории глухих и слабослышащих людей к единому банку жестов, что будет способствовать унификации жестового языка и смягчит проблему диалектов, и с разработки программных продуктов, обеспечивающих использование современных мультимедийных технологий для представления перевода в дактильной и жестовой форме.

7. Создание интернет портала «Сурдосервер»

По результатам проведенных исследований было признано целесообразным создать интернет портал, ориентированный на людей с нарушениями слуха и содержащий ресурсы по дактильным азбукам и жестовым языкам. Было признано, что наиболее актуальной и первоочередной задачей является создание базы видео файлов жестового языка и дактильной азбуки и обеспечение онлайн доступа к ним. До сих пор в сети не было этих ресурсов для русского языка.

Видео файлы предоставлены Всероссийским обществом глухих (ВОГ). Были произведены эксперименты по применению компьютерного распознавания речи для ввода слова или словосочетания. Введенное голосом слово преобразовывалось распознавателем в текст, и по тексту производился поиск в базе данных соответствующего видео файла. По результатам эксперимента было признано нецелесообразным использование распознавания речи для ввода слов и фраз с целью перевода их на жестовый язык и предложена система текстовых вводов. Такое решение было принято в связи с тем, что речь людей с нарушением слуха отличается плохой артикуляцией и плохо поддается компьютерному распознаванию.

Важной задачей также было признано ознакомление пользователей с жестовыми и дактильными языками мира, в том числе с международной дактильной азбукой и с международным жестовым языком. Такая возможность значительно расширяет кругозор людей с ослабленным слухом, снижает у них ощущение изолированности, оторванности от общества, пробуждает интерес к другим национальным культурам, и в конечном счете способствует их социальной реабилитации и раскрепощению.

В настоящее время «Сурдосервер» предоставляет пользователям следующие возможности: поиск и воспроизведение жестов по алфавиту, поиск и воспроизведение жестов по темам, русская дактильная азбука, дактильные азбуки более чем 30 стран мира, ссылки на сайты жестовых языков мира, добавление новых тем и пополнение банка жестов, обратная связь. «Сурдосервер» размещен в интернете и находится в режиме опытной эксплуатации. Адрес для доступа к нему http://surdoserver.ru/.

8. Архитектура web-приложения

Приложение основывается на PHP фреймворке Zend Framework 1.9.4 и использует архитектуру MVC (model-view-controller, модель-вид-контроллер), которая позволяет отделять логику работы с базой данных от логики внутренней работы приложения и от внешнего вида (представления) приложения.

Приложение состоит из нескольких модулей:

- Default - модуль по умолчанию - отвечает за первоначальную загрузку окружения приложения, подготовку вывода шаблона вида приложения в открытой части;
- Admin - модуль закрытой административной части приложения –- отвечает за систему аутентификации пользователей для доступа к административной части, также предоставляет интерфейс для административных частей других модулей. Содержит:
  - Контроллер аутентификации пользователей.
  - Программный интерфейс для работы административных панелей других модулей.
  - Шаблон административной части сайта.
- Surdo - модуль для работы со словами и категориями. Состоит из двух частей:
  - Открытая часть: отображение информации по категории, отображение информации по словам, начинающимся на определенную букву, отображение слова и всей сопроводительной информации по нему (видео-ролик, перевод и прочее).
  - Закрытая (административная) часть:
    - Интерфейс для работы с категориями: отображение списка категорий, добавление, удаление, редактирование категорий, добавление, удаление слов из категории, интерфейс множественного добавления слов в категорию.
    - Интерфейс для работы со словами: отображение списка слов, добавление, удаление, редактирование слова, интерфейс множественной загрузки слов на сервер.
- Dactyl - модуль для работы с дактильными азбуками:

  - Открытая часть: отображение списка имеющихся дактильных азбук, отображение определенной дактильной азбуки.
  - Закрытая (административная) часть: отображение списка имеющихся дактильных азбук, добавление, удаление, редактирование дактильной азбуки.
- Pages -модуль для работы с контентными страницами (такие страницы как «О проекте», «Другие жестовые языки» и другие):
  - Открытая часть: отображение списка контентных страниц, отображение определенной контентной страницы.
  - Закрытая (административная) часть: отображение списка имеющихся страниц, добавление, удаление, редактирование определенной страницы.

9. Технологическая и техническая платформа

Технологическая платформа сервера состоит из операционной системы - Ubuntu Server (linux) 8.04 LTS, веб-сервера - Apache 2.2, языка программирования - PHP 5.2.3, фреймворк - Zend Framework 1.9.4 и сервера баз данных - MySQL 5.0.14.

Сервер является VPS (Virtual Private Server, частный виртуальный сервер), арендуемым у компании «VPS.NET» (Великобритания, Лондон). «VPS.NET» реализует технологию «облачных вычислений», которая позволяет в короткое время (меньше минуты) расширять вычислительные способности арендуемого сервера до необходимых объемов, тем самым обеспечивая безотказную работу приложения при возрастании нагрузок.

Минимальные характеристики сервера: центральный процессор - 400 МГц, ОЗУ - 256 Мб, объем жесткого диска - 10 Гб, разрешенный трафик за месяц - 250 Гб.

10. Заключение

В нашей стране ощущается большой недостаток ресурсов для обучения и реабилитации людей с нарушениями слуха. Не хватает детских садов и школ для обучения глухих детей, лишь несколько ВУЗов в стране разрешают прием студентов с нарушениями слуха, глухим людям трудно получить среднее профессиональное образование. В то же время современные технологии и интернет предоставляют новые возможности, которые очень подходят для расширения познаний и кругозора глухих людей. «Сурдосервер» хоть и не заменит сурдопедагогов в их деятельности, но все же позволит всем желающим укрепить свои знания жестового языка путем самостоятельных занятий, поможет родителям глухих детей освоить дактиль и жесты. Он может использоваться также работниками социальных и других служб в изучении языка глухих.

Планируется расширить функции «Сурдосервера» в следующих направлениях: перевод русских слов и выражений на русский дактильный и жестовый языки, перевод русских слов и выражений на международный дактильный и жестовый языки, расширение возможностей переводчика, создание, популяризация и унификация новых жестов, разработка письменной формы русского жестового языка, создание методики обучения жестовому языку, раздел для детей.

Литература

1. Происхождение языка жестов. [Электронный ресурс]. – Режим доступа: http://psychologiya.ucoz.ru
2. Зайцева Г.Л. Дактилология. Жестовая речь: Учеб. для ВУЗов./Зайцева. – М.: «ВАЛДОС», 2004. [Электронный ресурс]. – Режим доступа: http://www.pedlib.ru/Books/l/0275
3. Воскресенский А.Л. Жестовая речь – язык или знаковая система? [Электронный ресурс]. – Режим доступа: http://www.dialog-21.ru/Archive/2004/Voskresenskij.htm
4. Воскресенский А.Л., Гуленко И.Е., Хахалин Г.К. Разработка средств перевода с естественного языка на жестовый язык глухих. [Электронный ресурс]. – Режим доступа: http://www.dom3.deria.ru/article/~pr/info_articles/5

С.В. Андреева

Лингвистические черты в зеркале систем автоматического понимания текста

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

Автоматическая обработка естественного языка (более точное выражение – обработка/анализ *текстов* на естественном языке) в нашей стране имеет полувековую историю. В 1955 году был проведён первый эксперимент по автоматическому переводу с одного языка на другой и начала складываться теория машинного перевода [1 - 8].

В последнее время на основе опыта работы систем машинного перевода разрабатываются крупные компьютерно-лингвистические приложения, в частности полифункциональный лингвистический процессор ЭТАП-3. В коллективной монографии известных лингвистов Ю.Д. Апресяна, И.М. Богуславского, Л.Л. Иомдина, В.З. Санникова «Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря» убедительно показано, что в синтаксическом анализаторе системы машинного перевода ЭТАП-3 «неизбежно остаются белые пятна и лакуны. Многие лингвистические факты, в том числе критически существенные для работы анализатора, никогда не попадали в поле зрения учёных просто потому, что они не имели возможности в массовом порядке оперировать материалом неверного или неожиданного синтаксического разбора предложений. Именно такой материал в изобилии поставляет развитая система автоматической обработки текстов, а работа с этим материалом позволяет выявить лакуны научного описания и устранить их» [9].

На основе анализа «уроков» функционирования формальной модели синтаксиса в системе машинного перевода с русского языка на английский авторами ставится и решается проблема пересмотра всей синтаксической модели русского языка в его взаимодействии с лексико-семантическим наполнением компонентов этой модели. Показательно, что этот пересмотр сделан на основе осмысления фактов автоматического перевода.

Авторы коллективной монографии приходят к неожиданному и своевременному выводу: «компьютерный анализ текстов является тем экспериментальным полигоном, на котором можно проверять не только отдельные грамматические и иные правила, но и целостную лингвистическую модель языка, т.е. адекватность лингвистической теории» [9].

С другой стороны, и создатели «интеллектуальных» систем, и лингвисты констатируют, что естественные тексты плохо поддаются формализации, и поэтому во многих действующих подходах преобладают эмпирические решения. Действительно, невозможно передать формальными средствами все языковые факты, особенно неоднозначные. Следовательно, возможности машинного перевода ограничены. Лингвистическое экспериментирование с участием систем автоматической обработки естественного языка «становится полноправным и весьма перспективным научным методом, а сами такие системы приобретают способность к (само) обучению. По результатам широкого лингвистического эксперимента, проводимого с помощью системы ЭТАП-3, делается попытка определить границы возможного в машинном переводе и – шире – вообще в автоматической обработке естественного языка» [9].

Машинный перевод с одного естественного языка на другой считают средой для создания систем автоматического понимания текста, относят к классу систем искусственного интеллекта. Однако при решении фундаментальной проблемы человеко-машинного диалога встаёт задача моделирования речевой деятельности. Как известно, значительную долю информации при общении между собой люди получают посредствам слуха. Однако большинство компьютерных систем, предназначенных для взаимодействия человека с

вычислительной машиной, пока недостаточно использует речевую информацию. Поэтому развитие технологий распознавания речи и многомодальных интерфейсов очень важно для создания новых интеллектуальных систем.

В этом аспекте представляются перспективными исследования речи на основе междисциплинарного подхода. Нельзя не отметить в связи с этим разработку методики подготовки комплекса баз данных, необходимых для работы системы распознавания речи, в том числе на уровне акустики, лексики и синтаксиса [10]. Процесс «взаимообогащения» лингвистической теории и «интеллектуальных» систем продолжается.

Литература

1. Мельчук И.А. Автоматический синтаксический анализ. – Новосибирск, 1964.
2. Мельчук И.А., Равич Р.Д. Автоматический перевод (1949 – 1963): Критико-библиографический справочник. – М., 1967.
3. Автоматический перевод: Сб. статей / Под ред. О.С. Кулагиной и И.А. Мельчука. – М., 1971.
4. Кулагина О.С. Исследования по машинному переводу. – М., 1979.
5. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы ЭТАП-2. – М., 1989.
6. Искусственный интеллект: Справочник: В 3 кн. – М., 1990.
7. Шаляпина З.М. Автоматический перевод: эволюция и современные тенденции // Вопросы языкознания. – 1996. – №2.
8. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы: учеб. пособие. – М., 2006.
9. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л., Санников В.З. Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря / Отв. ред. Ю.Д. Апресян. – М., 2010.
10. Ронжин А.А. Методы и программные средства многоканальной дистанционной обработки речи и их применение в интерактивных многомодальных приложениях. Автореф. дисс. … докт. техн. наук. – СПб., 2010.

Н.Б Вольская, С.Воробьёва

Нарушение линии деклинации как способ реализации акцентного выделения

Санкт-Петербургский университет,
Санкт-Петербург, Россия,
volni@phonetics.pu.ru

1. Введение

Под деклинацией, как известно, понимают свойство частоты основного тона постепенно понижаться от начала к концу высказывания [T'Hart et al 1990, Pierrehumbert 1980, Odé 1989]. Само явление впервые упоминается у К. Пайка (см. [Pike 1945]), но под другим названием; термин же деклинация (declination) был введен в обиход Дж. Т'Хартом, Р. Колье и А. Коэном [T'Hart et al 1990]. Многие исследователи рассматривают деклинацию как явление, претендующее на статус языковой универсалии, с большей или меньшей категоричностью (см. [Pierrehumbert 1980, Ladd 1998, Cruttenden 2000, Кодзасов, Кривнова 2001, Вольская 2000] и др.). Однако факт реальности деклинации подтверждается, в частности, экспериментами с синтезированной речью: фразы, синтезированные без учёта деклинации, воспринимаются как неестественные [T'Hart et al. 1990]. С другой стороны, есть и весомые аргументы против универсальности этого явления. Во-первых, наличие или отсутствие деклинации зависит от типа материала исследования. Так, если мы работаем с отдельными фразами, то деклинация, скорее всего, будет присутствовать. Это можно объяснить, например, ранним планированием говорящим конечного нисходящего тона обозначающего завершение высказывания. В естественной речи — связном тексте или спонтанном диалоге — количество фраз, реализованных с деклинацией, резко уменьшается [Cruttenden 2000: 121; Umeda 1981, 1982, цит. по Вольская 2000].

Принимая во внимание приведенную выше аргументацию, изучение явления деклинации представляется целесообразным по ряду причин:

- по экспериментальным данным, деклинация перцептивно значима: если синтезировать фразу с объективно равными пиками частоты основного тона, то слушатель будет воспринимать последующий пик как более высокий по отношению к предыдущему [T'Hart et al 1990: 142]. Эксперименты показали, что аудиторы при оценке сравнительной высоты тона ударных слогов в высказывании делают поправку на ожидаемую деклинацию [Pierrehumbert 1979];
- как показали данные эксперимента, проведенного на материале русского языка, деклинация играет существенную роль при оформлении различных коммуникативных типов (например, при противопоставлении вопросительного и повествовательного высказываний) [Вольская 2000];
- деклинация и определенные явления, с ней связанные, могут играть важную роль в синтагматическом членении фраз ([T'Hart et al 1990, Odé 1989, Collier et al 1993, Nagano-Madsen et al 1995, Вольская 2000] и др.).

Распространенной точкой зрения на причину возникновения деклинации в речи является отсылка к физиологическим ограничениям. Она обусловлена уменьшением величины подсвязочного давления [T'Hart et al 1990] и «связана с ослаблением дыхательного усилия ввиду расхода воздуха» [Кодзасов, Кривнова 2001: 386]. Чисто физиологическая трактовка подтверждается уже упомянутыми в предыдущих разделах наблюдениями о наличии деклинации в детском плаче, которые еще не умеют говорить, и у приматов. Косвенно с этими утверждениями связана теория существования

прототипической модели изменений частоты основного тона, которой обусловлено существование универсалий; в качестве основного контура сторонниками теории предлагается восходяще-нисходящая конструкция [Vaissière 1995, 2004].

Однако если деклинация перцептивно значима, можно предположить, что она является лингвистическим явлением и сознательно контролируется говорящим. В пользу языковой, а не физиологической природы деклинации свидетельствует сравнительное изучение особенностей вокализаций детей, которые еще не умеют говорить, в разных ситуациях: когда дети опробуют свои голосовые возможности, деклинация не наблюдается, в то время как при взаимодействии ребенка со взрослым, когда он еще не формирует предложения из отдельных слов, но коммуникативный тип его высказываний вполне угадываем, деклинация присутствует [Konopczynski et al. 1997].

Изучая явление деклинации, голландские ученые (IPO) пришли к выводу о том, что если в глобальном смысле падение частоты основного тона действительно от сознательной воли говорящего не зависит, то есть более локальное явление, которое можно считать проявлением контроля говорящего [T'Hart et al 1990: 140]. Речь идет о восстановлении линии деклинации (declination reset) на границе синтагм, как одном из трех наиболее часто используемых средств маркирования синтагматических границ, наряду с мелодическим переломом и паузами [T'Hart et al 1990: 125]. И если частотность использования именно это средства может варьироваться в зависимости от предпочтений диктора, то на восприятие слушателями выбор средства не влияет: какое бы средство или сочетание средств из этих трех ни было бы использовано, аудиторы одинаково хорошо отмечали границы между единицами [Collier et al. 1993].

Следует отметить, что тип синтагматической границы определяет интервал частоты основного тона между предграничным значением и точкой восстановления линии деклинации: так, увеличение от 40% до 155% характерно для границы предложения, в то время как на стыке синтагм внутри предложения это увеличение обычно не превышает 40% [Hata, Hasegawa 1992]. Результаты экспериментов с синтезом речи показывают, что стимулы, синтезированные без восстановления линии деклинации, по результатам перцептивного анализа были признаны неестественными.

Таким образом, такая модель — общая деклинация и восстановление линии деклинации на границах синтагм — является ожидаемой слушающим и влияет на восприятие речи.

Нарушение модели деклинации может использоваться говорящим как средство передачи акцентного выделения, то есть отражать и смысловую выделенность. В данной статье анализируются именно такие случаи акцентного выделения, обнаруженные в выборке из звукового корпуса русского языка CORPRES.

Акцентное выделение — это обозначение «активной для восприятия выделенности просодическими средствами какого-либо слова во фразе» [Николаева 2004: 3]. Понятия фразового ударения и акцентного выделения разноплановы. Фразовое ударение выполняет делимитативную функцию и функцию оформления коммуникативного типа высказывания — без фразового ударения фраза «не имеет цельнооформленности» [Николаева 2004: 9]. Акцентное же выделение этими функциями не обладает и никак не закреплено позиционно. Это значит, что а) они могут сосуществовать в одной фразе; б) в одной фразе может быть несколько акцентно выделенных слов; в) акцентное выделение может накладываться на фразовое ударение [Николаева 2004: 7-11].

О.Ф. Кривнова при исследовании различных мелодических акцентов во фразе и соответствующих акцентных контуров фраз отмечала, что исследования изменений тона только терминальной части высказывания недостаточно: «...можно предположить, что и в русских фразах (особенно повествовательных) может происходить фонетическое подчеркивание слов, создающее более выразительную акцентировку фразы, которую нельзя описать просто как контраст одного выделенного слова (носителя фразового ударения) и всех остальных слов как равновыделенных» [Кривнова 1978: 124]. Этой же

идеей руководствовались эксперты и при разметке звукового корпуса русского языка CORPRES, материалы которого использовались в настоящей работе. В рамках синтагмы выделялся интонационный центр — носитель синтагматического ударения, обозначаемый номером приписываемой ему интонационной конструкции. Кроме этого, эксперты в синтагме могли указывать и дополнительные акценты, отличные от интонационного центра, которые отмечались знаком [+]. Именно они и анализировались в данном исследовании.

Как критерии для описания акцентного выделения предлагается использовать две его характерные особенности:

- «акцентное выделение резко выделяет слово по сравнению с другими его соседями по высказыванию»;
- «акцентное выделение создает вокруг высказывания, в которое оно входит, определенный коммуникативно-текстовый и достаточно объективный для данного социума ореол» [Николаева 2004: 5].

Нас в данной работе интересует сам факт выделенности относительно других элементов фразы, отмеченной (и воспринятой) экспертами при разметке звукового материала, и интонационные способы реализации этой выделенности.

Акцентное выделение: просодические средства. Известно, что просодические параметры могут сочетаться для передачи синтагматического (или логического) ударения, причем нейтрализация одного из акустических средств компенсируется действием других [Галеева 1968; Кривнова 1978: 120; Светозарова 1982: 66-67; Vaissière 2004]. Однако относительно изученным в этом вопросе представляется лишь мелодический компонент, в то время как интенсивность, длительность, спектральные характеристики и т. д. рассматриваются значительно реже и менее системно. Некоторые сведения об особенностях реализации акцентного выделения представлены у С.В. Кодзасова [Кодзасов 2001] в разделе, посвященном описанию фонетической природы выделительных акцентов. Он пишет, что можно предположить, что легкость, с какой мы опознаем акцентно выделенные слова, «объясняется простым механизмом – совмещением перелома тона с ударением» [Кодзасов 2001]. Тем не менее, часты случаи, когда перелом тона отсутствует, и в то же время выделенность так же легко воспринимается, поэтому, по всей видимости, это недостаточное объяснение.

По наблюдениям С.В. Кодзасова, интенсивность также не имеет принципиального значения: «В акустическом сигнале не обнаруживается повышенной акустической интенсивности гласных, несущих выделительные акценты» [Кодзасов 2001].

Тем не менее и сам С.В. Кодзасов заключает, что «артикуляционные и акустические корреляты выделительного акцента требуют дальнейшего изучения» [Кодзасов 2001].

Роль мелодической составляющей в реализации различных типов реализации акцентов в зависимости от структуры предложения (простое, сложное, бессоюзное и т. д). рассматривается О.Ф. Кривновой в работе «Об акцентной функции мелодики» [Кривнова 1978]. Описание мелодических особенностей реализации акцентного выделения в выборке из звукового корпуса русского языка CORPRES можно найти в работе Качковской Т.В. [Качковская 2009].

В настоящем исследовании исследуется частный случай выделенности: реализация акцентного выделения за счет тонального контраста, ведущего к нарушению линии деклинации.

2. Материал и методика

Для анализа были выбраны фрагменты двух текстов, один на политическую, другой на техническую (компьютерную) темы — по 400 фраз каждого текста. В записи экспериментального материала участвовали 4 диктора: мужчины и женщины среднего возраста: профессиональные дикторы (GTE), актеры (STE, ATD) и преподаватели (OTD). Объем экспериментального материала составил 1600 фраз. Далее из этого материала

отбирались те фразы, которые, согласно экспертной разметке, содержали в себе акцентное выделение, которое в интонационной разметке, принятой в корпусе CORPRES [Вольская, Скрелин 2009] отмечалось знаком [+] , например:

[грустным выглядел лишь [+]открывший торжественную [12]церемонию /p3/ экс-[02]спикер /p3/ владимир [02]литвин /p3/ который [+]сменил президиум верховной [12]рады /p3/ на президиум национальной академии [02]наук украины /p3/ [12]недавно /p2/ его избрали вице-президентом [01c]нан /p5/]

Следующим шагом стал отбор тех фраз, в которых акцентное выделение находилось в синтагме, оформленной интонацией завершенности, т.е. рассматривались такие синтагмы, в которых движение тона было нисходящим, независимо от коммуникативного типа высказывания, и АВ реализовано как результат отклонения от линии деклинации.

В основе исследования лежала гипотеза о том, что «в синтагме, где интонационный центр оформлен нисходящим тоном, акцентное выделение реализуется подъемом тона — дополнительной вершиной в синтагме» [Вольская, Скрелин 2009].

3. Частотность случаев акцентного выделения в выборке

В таблице 1 представлено отношение фраз с акцентным выделением к общему количеству рассмотренных фраз.

Таблица 1. Соотношение количества случаев АВ с общим объемом выборки для каждого диктора.

	Технический текст		**Политический текст**	
	Диктор 1 (GTE)	Диктор 2 (STE)	Диктор 3 (ATD)	Диктор 4 (OTD
Всего	400	400	400	400
Случаев АВ	177	130	287	71
В процентах	**44,25%**	**32,50%**	**71,75%**	**17,75%**

Количество фраз, в которых АВ реализовано в синтагмах, оформленных нисходящей интонацией, представлено в Таблице 2.

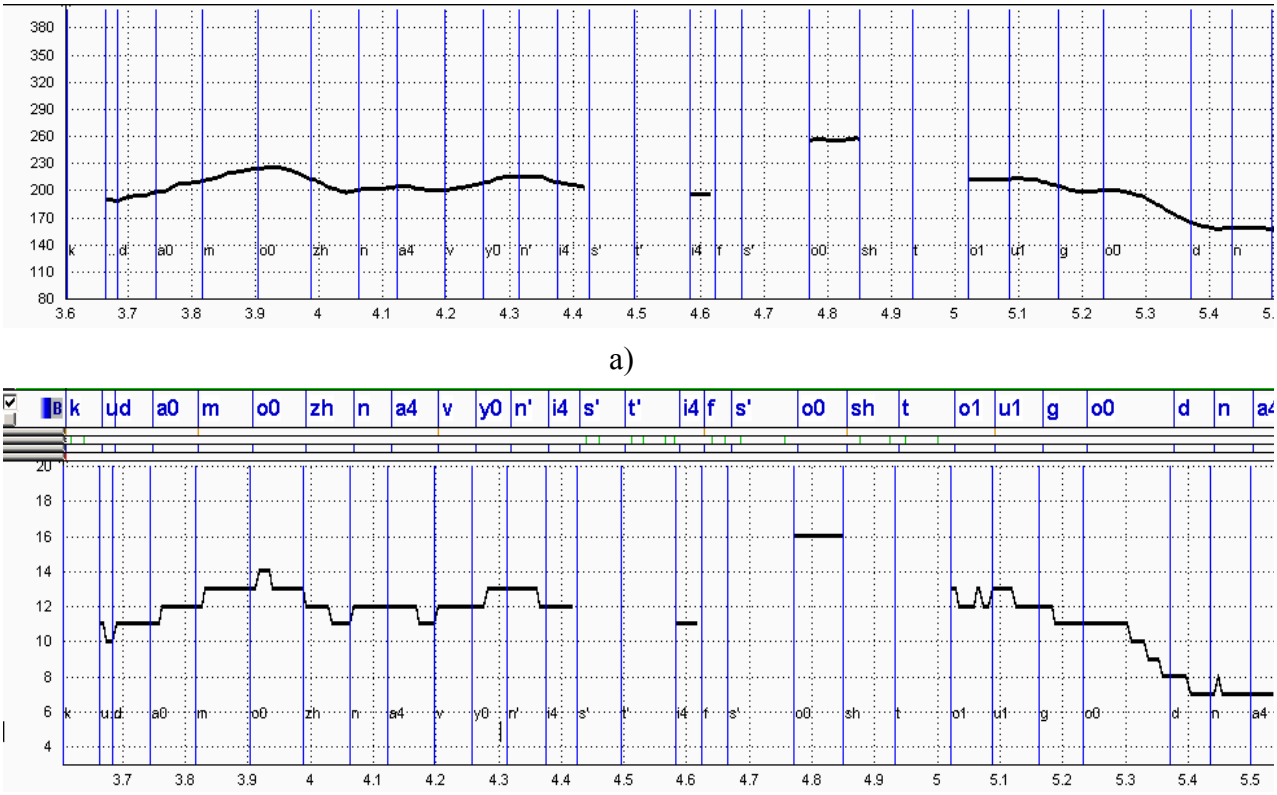
Таблица 2. Соотношение фраз «АВ+нисх» с общим количеством фраз с АВ

	Технический текст		**Политический текст**	
	Диктор 1 (GTE)	Диктор 2 (STE)	Диктор 3 (ATD)	Диктор 4 (OTD
Случаев АВ	177	130	287	71
АВ+нисх	27	33	33	12
АВ+ нисх (%)	**15,25%**	**25,39%**	**11,50%**	**16,90%**

В основе анализа лежала гипотеза о том, что нарушение ожидаемой линии деклинации может выполнять функцию выделения элементов речи. В соответствии с этим предполагалось, что главным средством оформления акцентного выделения будет подъём частоты основного тона на ударном слоге выделяемого слога, превышающий или равный по значению уровню частоты основного тона ударного слога предыдущего слова (что приводит к существенному нарушению линии деклинации). Результаты экспериментов показывают, что в сознании носителя языка существует определенная модель деклинации, которую он ожидает услышать в зависимости от длины высказывания и его мелодического оформления. Поэтому при восприятии речи слушающий делает поправку на эту модель и определенным образом воспринимает и нарушения своих ожиданий [Pierrehumbert 1979].

Из 44 рассмотренных случаев 29 (т.е. 66%) подтверждают это предположение. Далее случаи подразделяются на группы в зависимости от обнаруженных акустических закономерностей.

Группа 1. Из указанных 29 случаев в 18-ти значение частоты основного тона ударного слога акцентно выделенного слова превышает значение частоты основного тона ударного слога предыдущего слова (Рис. 1). Диапазон между ударным слогом выделяемого слова и ударным слогом предыдущего слова может быть различным: от 10-15 до практически 50 Гц. Это соответствует превышению в один-два полутона, в предельных случаях разница достигает 4 полутонов, хотя в отдельных ситуациях возможно и равенство значений в полутонах.



а)

б)

Рис. 1 График изменения ЧОТ фрагмента синтагмы : *куда можно вынести [+]всё что [01a]угодно a) в герцах b) в полутонах* .

Значение частоты основного тона ударного слога «все» составляет 255 Гц и 12.5 полутонов (средние значения частоты основного тона в слоге), "фонового" предшествующего ударного слога в слове «вынести» - 212 Гц и 16 полутонов (средние значения частоты основного тона в слоге).

К этим случаям также относятся ситуации, где тон в рамках ударного слога выделяемого слова сильно поднимается (на 20-30 Гц), так что верхняя точка подъема также находится выше значения частоты основного тона предшествующего ударного слога. В полутонах это восходящее движение осуществляется в диапазоне 3-х или 4-х полутонов (см. Рис. 2 a), b)). Подъем частоты основного тона от "фонового" ударного слога слова «которая» до выделенного "очень" — от 4 до 8 полутонов.

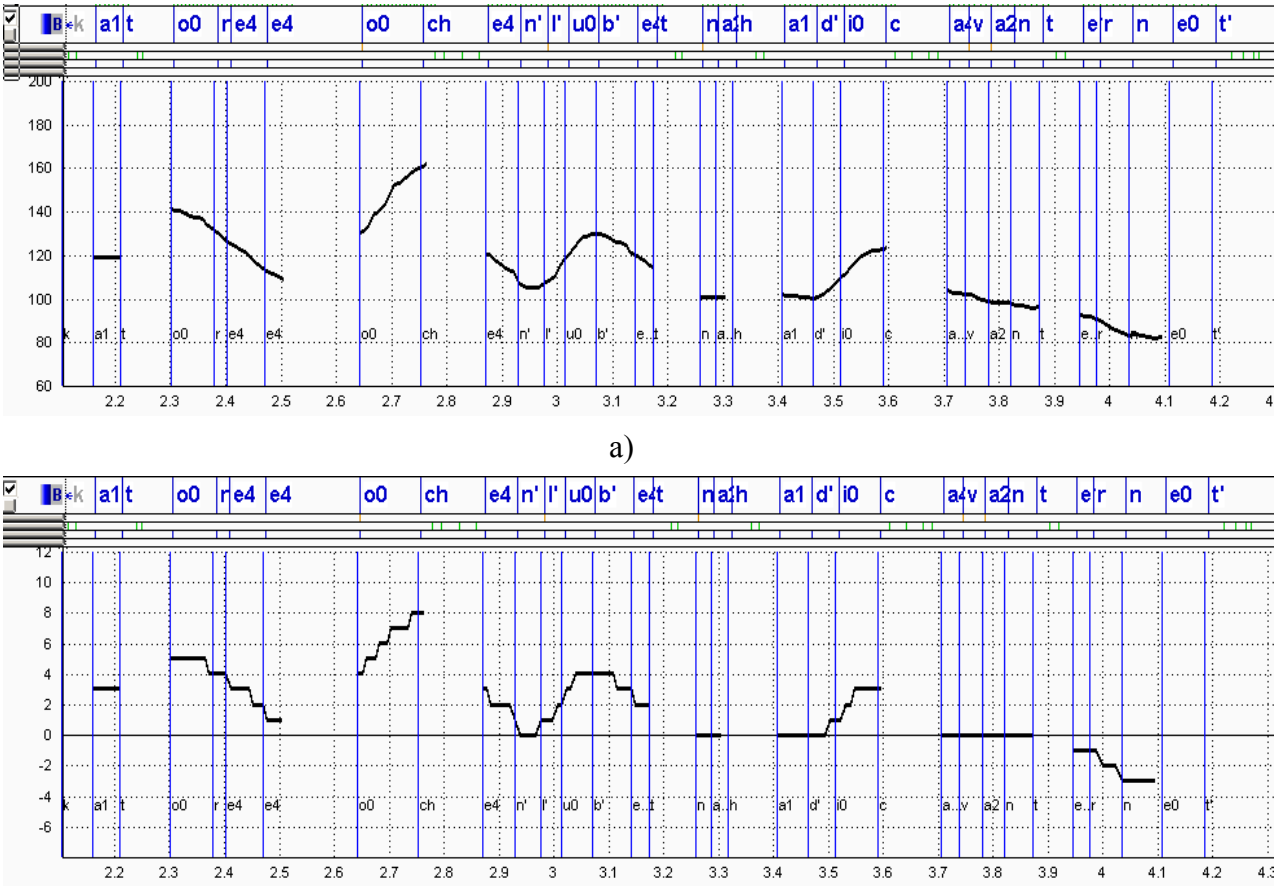


Рис. 2. График изменения ЧОТ синтагмы *которая [+]очень любит находиться в [01a]интернете* a) в герцах и b) в полутонах.

Группа 2. В оставшихся 11 случаях (из 29 случаев с очевидным нарушением линии деклинации) значение частоты основного тона ударного слога выделяемого слова не превышает тональный уровень предыдущего ударного слога, а достигает его значения (Рис. 3 a) и b)). Разница в тональном уровне "фонового" ударного слога слова "*себе*" и носителя акцентного выделения ударного слога слова "колоссальную" всего 0,5 полутона (среднее значение частоты основного тона ударного слога в слове «себе» составляет 158 Гц (7.5 полутонов), среднее значение частоты основного тона ударного слога в «колоссальную» равно 162.5 Гц (8 полутонов).

Группа 3. В оставшихся 15 случаях не наблюдается столь явных признаков выделенности за счет движения частоты основного тона. Чем тогда объясняется восприятие слов как акцентно выделенных — вопрос для дальнейшего изучения с точки зрения других акустических параметров (интенсивности, длительности, спектральных характеристик и т. д.): известно, что просодические параметры могут сочетаться для передачи синтагматического (или логического) ударения, причем нейтрализация одного из акустических средств компенсируется действием других [Галеева 1968; Кривнова 1978: 120; Светозарова 1982: 66-67; Vaissière 2004].

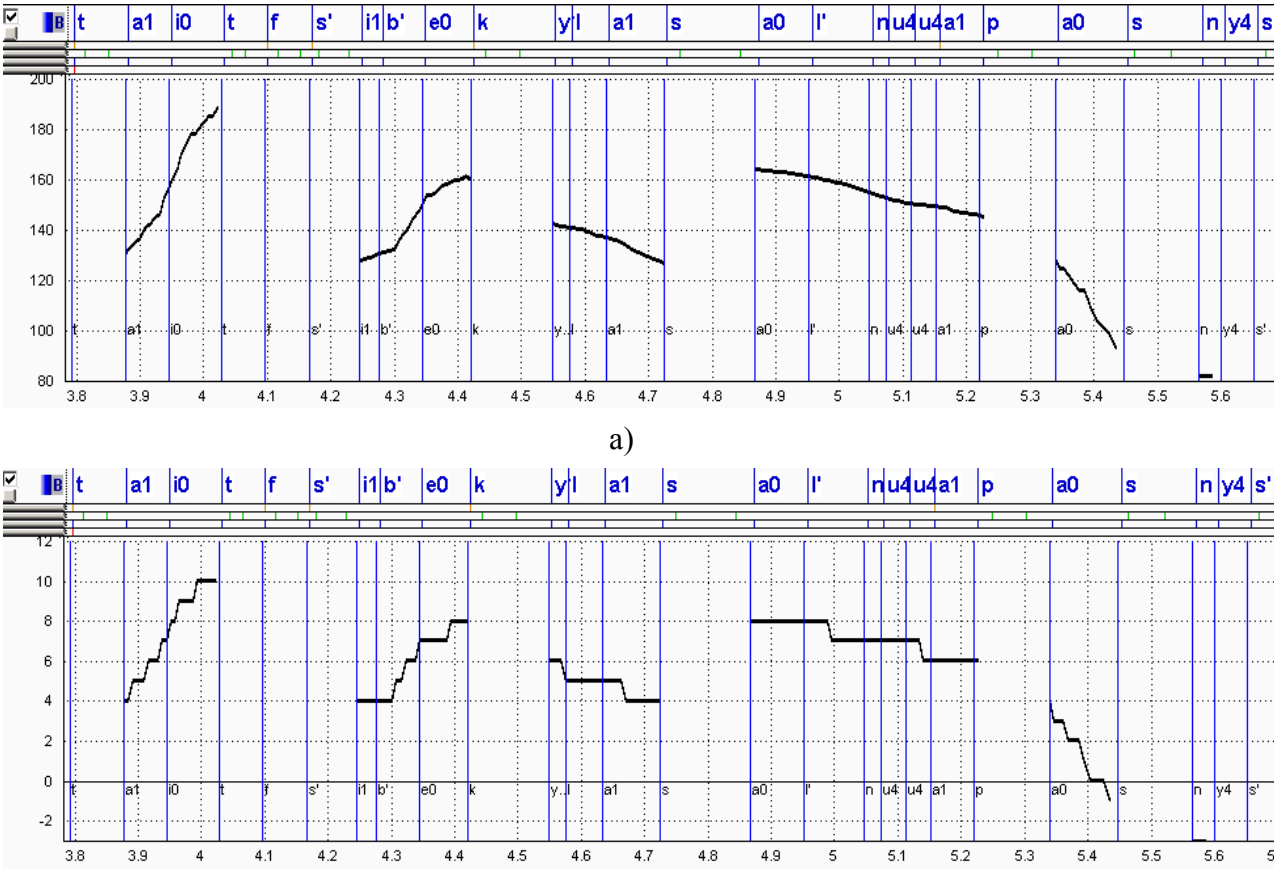


Рис. 3. График изменения ЧОТ синтагмы (Гц): *таит в себе [+]колоссальную [02]опасность* а) в герцах b) в полутонах.

4. Заключение

Подводя итоги, можно выделить следующие тенденции:

- значительное большинство случаев акцентного выделения (66%) реализуется за счет нарушения линии деклинации: выбросом частоты основного тона ударного слога выделяемого слова до уровня, либо превосходящего значение частоты основного тона ударного слога предыдущего слова, либо равного ему.
- это объясняет не все случаи восприятия слова как акцентно выделенного, для их трактовки необходимо дальнейшее изучение на большем объеме материала, затрагивающее другие акустические признаки, которые потенциально могут способствовать восприятию выделенности.

Литература

1. Вольская Н.Б. Деклинация: к проблеме моделирования интонации для синтеза речи // Zmogus Kalbos Erdveje, pages 194 – 199. Kaunas, 2000
2. Вольская Н.Б., Скрелин П.А. Система интонационных моделей для автоматической интерпретации интонационного оформления высказывания: функциональные и перцептивные характеристики // Труды семинара АР 3-2009, Санкт-Петербург. С. 28 – 40
3. Галеева М.М. Элементы интонации и их взаимодействие в синтагмах повествовательного предложения в русском языке. Дисс. канд. филол. наук. 1968
4. Качковская Т.В. Способы реализации интонации выделенности в русском языке // Квалификационная работа. С-Пб, 2009
5. Кодзасов С.В. Законы фразовой акцентуации. // Просодический строй русской речи. М., 1996. С. 181 – 204

6. Кодзасов С.В. Комбинаторные методы в фонологии // Докторская диссертация МГУ. М., 2001
7. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М., 2001
8. Кривнова О.Ф. Об акцентной функции мелодики (на материале русского языка) // Интонация. Киев: Вища школа, 1978. С. 119 – 143
9. Николаева Т.М. Семантика акцентного выделения. М.: УРСС, 2004
10. Павлова А.В. Акцентная структура высказывания в ее связях с лексической семантикой. Автореферат дисс. Л.: ЛГУ, 1987
11. Павлова А.В. Значение просодической информации в лексикографическом толковании полисемии и омонимии // Труды международной конференции «Диалог-2008». Москва, 2008. С. 407 – 410
12. Светозарова Н.Д. Интонационная система русского языка. Л., 1982.
13. Babkin A., Krivnova O. F0 contour generation in TTS system for Russian language // International workshop 'Prosody 2000: speech recognition and synthesis. Poland, Krakov, 2000.
14. Konopczynski G., Santi S., Embarki M., Bertrand R. F0 declination line: new evidence fom infants and children's speech (9-36 months), a preliminary research // INT-1997, Greece, Athens. Pp. 199 – 202
15. Ladd D.R. Declination: a review and some hypotheses // Phonology yearbook, vol. 1, 1984. Pp. 53 – 74.
16. Ladd D. R. Intonational phonology. Cambridge: CUP, 1998
17. Lieberman P., Crelin E.S., Klatt ВюРю Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee" // American Anthropologist 74, 1972. Pp. 287 – 307
18. Nagano-Madsen Y., Bredvard-Jensen A.-C. Comparing F0 reset in two prosodically diverse languages (data from Eskimo and Yoruba) // ICPhS 1995, vol. 3, Stockholm, 1995. Pp. 652 – 655.
19. Odé C. Russian intonation: a perceptual description. Amsterdam: Rodopi, 1989
20. Pierrehumbert J. The perception of fundamental frequency declination // JASA 66(2) 1979. Pp. 363 – 369
21. Pierrehumbert J. The phonology and phonetics of English intonation. PhD thesis. 1980
22. Pike K. The intonation of American English. University of Michigan: 1945
23. Swerts M., Strangert E., Heldner M. F0 declination in read-aloud and spontaneous speech // ICSLP-1996, Philadelphia, PA, USA. Pp. 1501 – 1504
24. t'Hart J., Collier R., Cohen A. A perceptual study of intonation: an experimental-phonetic approach to speech melody. Cambridge: CUP, 1990
25. Vaissière J. Natural explanations for prosodic cross-languages similiraties // ICPhS 1995, vol. 4, Stockholm, 1995. Pp. 654 – 657
26. Vaissière J. Perception of intonation. // Ed. by D.B. Pisoni, R.E. Remez. Handbook of speech perception. Oxford, Blackwell: 2004. Pp. 236 – 263

Научное издание

ЧЕТВЕРТЫЙ МЕЖДИСЦИПЛИНАРНЫЙ СЕМИНАР
«АНАЛИЗ РАЗГОВОРНОЙ РУССКОЙ РЕЧИ» АР^3 - 2010

Труды семинара

Составитель: А.Л. Ронжин

Издание осуществлено за счет средств проекта «Разработка методов человеко-машинного взаимодействия и многомодальных пользовательских интерфейсов для интеллектуальных информационных систем» по Государственному контракту № П2360 в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы

Подписано в печать 08.06.10. Формат 60 x 84/8.
Печ. л. 7,5 Тираж 70 экз. Заказ № 320.

Редакционно-издательский центр ГУАП
190000, Санкт-Петербург, Б. Морская ул., д. 67.