

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>

Труды третьего междисциплинарного семинара
«Анализ разговорной русской речи» (АР³ - 2009). – СПб.: ГУАП, 2009. – 84 с.
ISBN 978-5-8088-0480-7

Издание представляет собой сборник докладов, сделанных на заседаниях третьего междисциплинарного семинара «Анализ разговорной русской речи» (АР³ - 2009), проходившего 26 – 27 августа 2009 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 978-5-8088-0480-7

© СПИИРАН, 2009
© Коллектив авторов, 2009
© ГУАП, 2009

Перцептивный словарь и автоматическое распознавание речи

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

В системах автоматического распознавания речи, как и в системе восприятия речи человеком, первичное перекодирование входного акустического сигнала в некоторую последовательность «символов», доступную для последующей обработки, базируется на результатах спектрального анализа исходного сигнала. При этом возникают сложности с интерпретацией тех участков сигнала, которые относятся к гласным звукам. Возникает проблема предварительной «нормализации» спектральной картины анализируемого речевого сигнала для компенсации хорошо известных различий, характерных для дикторов разного пола и возраста. Но даже речь дикторов одного пола и возраста характеризуется значительными вариациями частот формант даже при произнесении так называемых кардинальных гласных русского языка (рис. 1).

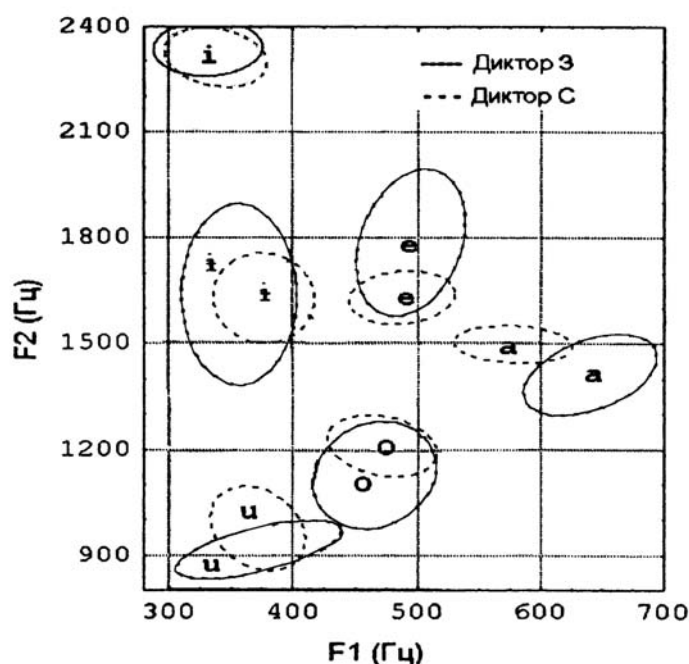


Рис. 1. Распределение частот формант гласных, произнесенных во фразе «Повтори [sVs] опять» (по: [2]).

Результаты же работы со спонтанной речью свидетельствуют, что качество гласных, характеризуемое их формантной структурой, зачастую совершенно не отвечает кодифицированным орфоэпическим нормам русского языка. Причем касается это не только безударных гласных, но и гласных в ударных слогах. Характерным примером такого рода отклонений может служить произнесение одним из участников беседы на «Эхо Москвы» во втором предударном слоге фонетического слова «в абсолютно» гласного, спектральные характеристики которого весьма близки к таковым для ударного гласного «о» (рис. 2), хотя в соответствии с классическими правилами русской транскрипции в этой позиции должен был звучать [ь].

Аналогичные отклонения от нормы можно наблюдать и в первом предударном и даже в ударном слогах: [нъ черевнъ] (*на царевне*) или [зд'ис'ис'ич'ас] (*здесь и сейчас*).

В этих условиях в системе автоматического распознавания или системе восприятия речи человеком для надежной работы с любым речевым сигналом должна быть предусмотрена подстройка под индивидуальные свойства любого диктора, причем мгновенно работающая подстройка, поскольку для стороннего слушателя адекватное понимание диалога двух разных дикторов не представляет никакого труда.

К этому надо добавить, что в спонтанной речи нередки случаи потери редуцированными словоформами ударных гласных и это не приводит к потере ее разборчивости, а в восстановлении «утрат» носитель языка в значительной степени опирается на более устойчивый к редукциям консонантный скелет словоформы и при этом важную роль играет контекст [3].

Таким образом, оказывается, что при восприятии спонтанной речи носитель языка способен восстанавливать обусловленные редукцией утраты в сегментном составе словоформ и в ряде случаев может обходиться без достоверной информации о фонемном качестве гласных и даже без самих гласных.

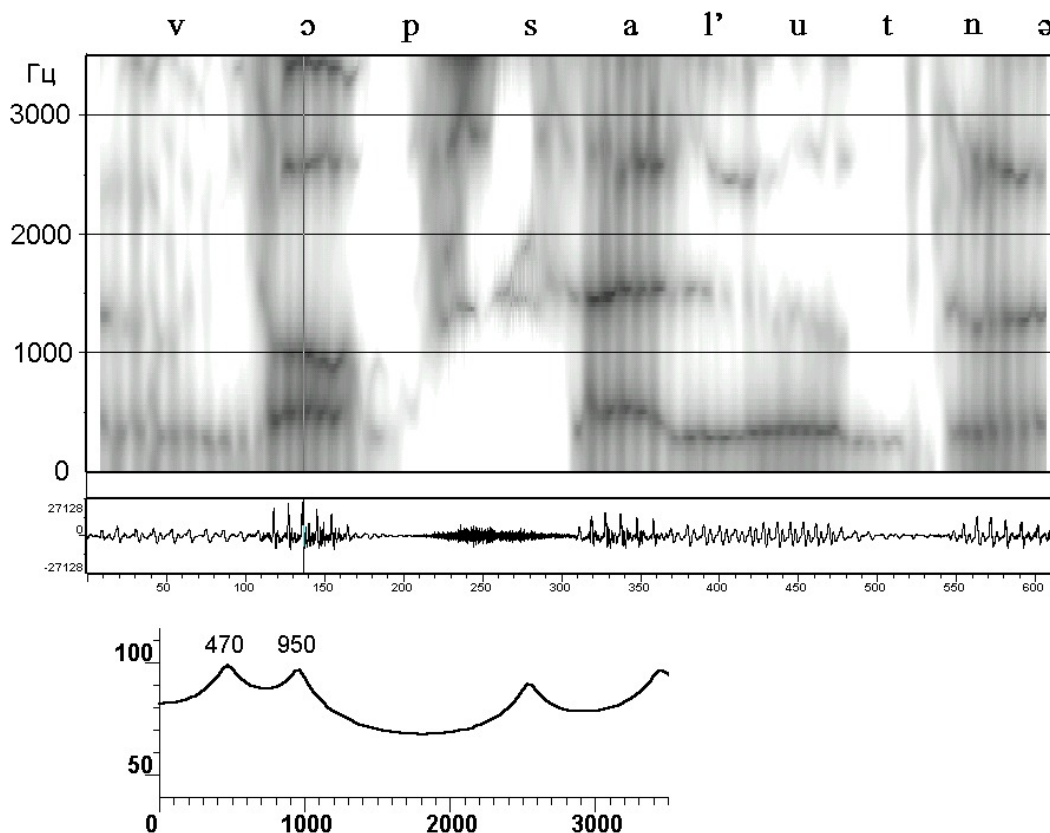


Рис. 2. Динамическая спектрограмма фонетического слова «в абсолютно» (вверху) и спектральный срез в средней части второго предударного гласного (внизу).

Предположим теперь, что используя при распознавании перцептивный словарь и учитывая для снятия многозначности правила разных уровней анализа (морфологического, синтаксического, семантического и т.д.), можно будет обойтись без идентификации фонемной (аллофонной) принадлежности гласных, а ограничиться только констатацией факта их наличия в распознаваемом сигнале и признаком «ударный/безударный», т.е. воспользоваться значительно менее вариативными свойствами консонантного скелета словоформ.

Чтобы оценить степень реалистичности такого подхода можно воспользоваться одним из текстов Корпуса русского литературного языка (www.nagusco.ru) и созданным на базе корпуса частотным словарем словоформ [1]. Очевидно, что в предполагаемой функциональной модели восприятия речи, как и в системах автоматического распознавания, придется иметь дело с акустико-фонетическим (аллофонным) представлением распознаваемого текста и с перцептивным словарем, элементы которого описаны той же системой символов. К сожалению, на сегодняшний день в нашем распоряжении нет ни того, ни другого, и все же проведенный эксперимент позволяет прояснить круг возможных проблем, даже с использованием в виде весьма приблизительного аналога перцептивного словаря словоформ в их орфографической записи в виде весьма приблизительного аналога перцептивного словаря.

Чтобы несколько приблизить параметры данной модели к предполагаемым характеристикам системы восприятия, процедура сопоставления словоформ текста с единицами словаря сопровождалась палатализацией согласных и уравниванием **тбся** – **тсся** в глагольных словоформах.

На рисунке 3 приведена часть текста из [4] с предварительным результатом «распознавания». Ударность гласного маркирована символом «+», а в угловых скобках приведены все имеющиеся в частотном словаре словоформы, совпадающие по количеству гласных и консонантному скелету со словоформой из распознаваемого текста. В круглых скобках приведены частоты словоформ в словаре, полученном на корпусе объемом около 1 млн. словоупотреблений.

Но+чь, мерца+ющие боло+тные
 <огоньки+(6) огоньке+(3)>,
 <венке+(2) венки+(3)>,
 <пе+сни(70) пе+сня(40) пи+сня(1) пе+сне(7) пе+сню(28)>,
 жу+ть,
 <ползу+щую(1) ползу+щая(1)>
 по
 <позвоно+чника(1) позвоно+чнику(1)>
 при
 <мы+сли(135) ма+сле(1)>
 <а(9795) о(3235) у(5496) и(35922)>
 <по+иске(4) по+иски(22)> [по+ьскъ] (26)
 цвету+щего па+поротника, -- вку+с те+рпкой го+голевской
 <зава+рки(4) зава+рке(1)> [зава+ркъ] (5)
 знако+м с
 <де+тство(23) де+тства(61)>, [де+тствъ] (84)
 приме+рно
 та+к<та+к(2795) то+к(14) ту+к(3)>
 же, как
 <шо+рах(1) шо+рох(8)> [шо+ръх] (9)
 кры+льев
 <де+моны(1) де+мона(2)> [де+монъ] (3)
 и+мени Вру+бея - Ле+рмонтова.

 <на(16067) нэ(5)>
 <са+мым(120) са+мом(89)> [са+мъм] (209)
 <де+ле(42) де+ли(2)>, [де+ль] (44)
 демони+ческого в
 <весё+лым(13) весё+лом(4)> [весё+лъм] (17)
 <пра+зднике(4) пра+здники(10)> [пра+зднькъ] (14)
 <Ива+на(42) Ива+ну(5) Ива+ны(1) Иво+на(2)>
 <Купа+лы(4) Купа+ла(5) Купа+лу(1)>
 <бы+ло(2452) ба+ла(4)>
 бы не
 <та+к(2795) то+к(14) ту+к(3)>
 уж
 <мно+го(635) мно+гу(1)>,
 <е+сли(2072) я+сли(3)>
 бы
 <вели+кую(13) вели+кое(23) вели+кая(32)>
 <ру+сскую(16) ру+сское(19) ру+сская(35)>
 <литерату+ру(14) литерату+ры(58) литерату+ра(20)>
 не расстара+лась.

Рис. 3. Пример предварительного результата «распознавания»

Всего в тексте было обработано 577 словоформ, из которых 361 (62,6%) получила однозначную интерпретацию (рис. 4): в частотном словаре была обнаружена единственная реализация, соответствующая консонантному скелету словоформы, числу и позиции

ударного слога. Сорок из оставшихся 216 словоформ не могли бы быть однозначно идентифицированы даже при идеальном сохранении фонемного качества гласных, поскольку различия реализовывались только в заударных слогах (см. примеры приблизительной транскрипции в квадратных скобках), а следовательно, в процедуру распознавания должны быть включены правила лингвистического анализа.

Полученный результат показывает также, что единицы перцептивного словаря должны снабжаться грамматическими описателями, используя которые и учитывая ближайший контекст, система может выбрать из списка кандидатов элемент с правильной падежной формой (*знако+м с [де+тствь]*), или являющийся нужной частью речи (*на са+мом [де+ль]*), или выбрать кандидата с нужным значением (*как [шо+рьх] крыльев*).

Этот же принцип может быть использован и при идентификации односложных знаменательных слов: *с Ива+нова <дня+(253) дни+(163) дне+(29) дню+(12)>* дает единственно возможный результат – «с Иванова дня», так как предлог «с» требует либо родительного, либо творительного падежа, а из имеющихся в словаре кандидатов только «дня» отвечает этому условию.

Однако не всегда разрешению неопределенности при выборе из нескольких кандидатов помогает ближайший контекст. Так выбор именительного падежа (*великая русская литература*) может быть сделан только после правильного распознавания глагола *расстаралась*. Очевидно, что для этого потребуется буфер памяти, в котором некоторое время будут храниться все кандидаты, а также какие-то правила сканирования буфера.

Из всего сказанного можно заключить следующее. Во-первых, идея отказаться от фонемной идентификации гласных при распознавании речи оказывается не совсем уж безрассудной, если единицы используемого перцептивного словаря снабжены необходимыми грамматическими и семантическими пометами, а в алгоритм распознавания включены правила использования этих помет. Во-вторых, подобного рода эксперименты с текстами и словарем в орфографической записи позволяют приблизиться к пониманию того, как может быть устроен перцептивный словарь, и сформулировать правила использования знаний грамматики и семантики в алгоритмах автоматического распознавания или моделях восприятия речи человеком.

Литература

1. Венцов А.В., Е.В. Грудева. Частотный словарь словоформ русского языка (проект). Череповец: Изд-во ЧГУ, 2008. 204 с.
2. Кузнецов В.Б. О принципах акустической классификации русских гласных. Язык и речь: проблемы и решения: Сборник научн. трудов к юбилею проф. Л.В. Златоустовой / Под ред. Г.Е. Кедровой и В.В. Потапова. М.: МАКСПресс, 2004. С. 100–116.
3. 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.
4. Купала весёлый, Купала любовный // Невское время.- 1998.- N 111(1753), 23 июня. <http://www.nvrem.dux.ru/>.

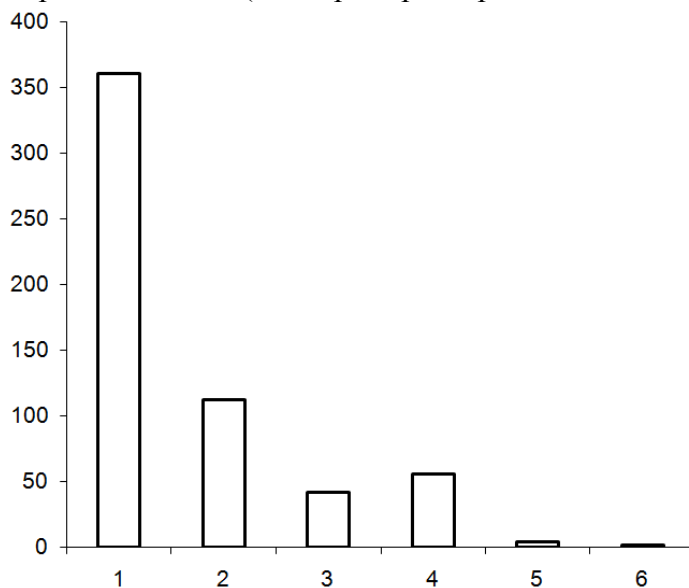


Рис. 4. Распределение числа слов текста по числу кандидатов в словаре