

УДК 004.93

*А.Л. Ронжин, А.А. Карпов, И.В. Ли*Санкт-Петербургский институт информатики и автоматизации РАН, Россия,
ronzhin@iias.spb.su, karpov@iias.spb.su, lee@iias.spb.su

Система автоматического распознавания русской речи SIRIUS*

В статье представлена разработанная в группе речевой информатики СПИИРАН система распознавания слитной русской речи SIRIUS. Особенностью данной системы является наличие в ней морфемного уровня представления языка и речи, что позволяет значительно сократить размер лексического распознаваемого словаря и скорость обработки. Описан процесс внедрения системы распознавания русской речи в области инфотелекоммуникаций для голосового доступа к интернет-версии электронного каталога «Желтые страницы Санкт-Петербурга» с целью создания автоматизированного call-центра для ответов на звонки абонентов. В статье представляются результаты тестирования системы как при работе с речью, записанной в офисных условиях, так и в условиях телефонного разговора.

Введение

Информационные технологии все больше и больше проникают в повседневную жизнь каждого человека. Технические устройства развиваются в сторону интеллектуализации и автоматизации предоставляемых ими сервисов и услуг. Особенно активно развиваются системы искусственного интеллекта, связанные с распознаванием образов, анализом изображений и речи. В последнее время особое внимание как исследователей, так и конечных пользователей уделяется разработке и применению автоматизированных систем, построенных с использованием систем распознавания речи. Речевые технологии находят все большее распространение в робототехнике, системах управления оборудованием, средствах телекоммуникаций.

Такие приложения уже являются реальностью в странах США и Европы. В России сейчас также предпринимаются некоторые попытки разработать и внедрить подобные системы и сервисы. Основной проблемой здесь является трудность применения технологий распознавания речи для русского языка из-за сложного механизма словообразования. В отличие от, скажем, английского языка русский язык обладает гораздо большей вариативностью на уровне словоформ. В связи с чем резко возрастает размер распознаваемого словаря и падает точность. Попытки наложения синтаксических ограничений на распознаваемые фразы приводят к тому, что ошибки на уровне окончаний слов или приставок приводят к ошибке распознавания всей произнесенной фразы. Кроме того, естественная речь содержит множество неточностей и отклонений, которые не могут быть учтены заранее. Неудача существующих на рынке подходов к распознаванию и пониманию речи состоит в том, что они не обладают

* Данные исследования проводятся при финансовой поддержке Правительства Санкт-Петербурга, Фонда содействия отечественной науке и Европейского Сообщества SIMILAR NoE FP6: IST-2002-507609, а также проекта INTAS № 04-77-7404.

достаточной робастностью алгоритмов распознавания и понимания речи к различным возможным искажениям сигнала на входе системы и собственным ошибкам обработки, а также их недостаточно высокой точности работы.

Для решения указанных проблем группой речевой информатики СПИИРАН был введен дополнительный уровень представления языка и речи – морфемный уровень и разработана оригинальная система распознавания русской речи SIRIUS (SPIIRAS Interface for Recognition and Integral Understanding of Speech) [1]. За счет разделения словоформ на морфемы словарь распознаваемых лексических единиц значительно сократился, так как в процессе словообразования часто используются одни и те же морфемы. На основе правил словообразования русского языка были разработаны базы данных различных типов морфем, а также методы автоматической обработки текстов. Разработанные базы данных морфем далее использовались для накопления статистики встречаемости морфем по текстовым корпусам, причем в ходе распознавания основное значение имеет степень согласованности корневых морфем. В результате такой обработки скорость распознавания и устойчивость к синтаксическим отклонениям в произнесенной фразе увеличились.

В процессе создания лексических баз данных использован базовый словарь Зализняка размером около 160 000 слов [2]. При распознавании слитной речи обеспечена дикторнезависимость по отношению к носителям русского языка, а также точность свыше 90 % на словаре около 2 000 слов конкретной предметной области. Кроме того, модель была опробована при распознавании речи через телефонный канал. Представлены предварительные результаты и методы, использованные для очистки речевого сигнала от шумов. Разработанные модули и базы данных собраны в единый аппаратно-программный комплекс для автоматического ввода и распознавания русской речи.

1 Особенности распознавания русской речи

На сегодняшний день активнее всего ведутся исследования и разработки систем автоматического распознавания английской речи, поэтому далее приведем несколько отличий русского и английского языков, а также ряд проблем, возникающих при распознавании русской речи.

В русском языке по сравнению с английским более сложная структура словообразования (из-за того, что русский язык является синтетическим языком), в результате чего необходимо использовать гораздо больший распознаваемый словарь, что значительно уменьшает точность распознавания и скорость распознавания. В передовых системах распознавания речи для английского языка (от фирм Microsoft, Scansoft, IBM) используется словарь свыше 100 тыс. слов. Для русского языка за счет наличия приставок, суффиксов и окончаний этот словарь возрастает на порядок и более. Грамматический словарь А.А. Зализняка содержит около 160 тысяч наиболее употребительных слов русского языка, при помощи специальной системы обозначений он позволяет построить все словоформы для выбранного слова. При развороте всех словарных статей получается около 3,7 млн словоформ. Кроме того, большинство словоформ одного и того же слова отличаются только в окончаниях, которые произносятся обычно не так четко, как начала слов. А ошибки в окончаниях при распознавании слов приводят к тому, что происходит ошибка в распознавании всей фразы из-за несогласованности слов.

Порядок слов в предложении русского языка не задан жестко, а может варьироваться без потери смысла предложения, в английском же языке используются жесткие грамматические конструкции. Это затрудняет создание статистических моделей языков на основе биграмм или N-gram, а также грамматик для русского языка и понижает их эффективность.

Различен фонетический алфавит языков. В международном фонетическом алфавите SAMPA для русского языка принято 44 фонемы: 38 фонем согласных звуков и 6 фонем гласных звуков. В американском варианте английского языка фонетический алфавит SAMPA насчитывает 41 фонему: 24 согласных и 17 гласных (включая целый ряд дифтонгов). Очевидно, что распознавание согласных звуков сложнее, чем гласных, из-за того, что они менее стабильны, чем гласные, и имеют гораздо меньшую длительность.

В русском языке в среднем более длинные слова, чем в английском языке, что замедляет процесс распознавания, так как необходимо анализировать более длинные участки сигнала.

Сложность распознавания русской речи связана также с тем, что Россия является многонациональной страной и каждая нация владеет, помимо русского языка, еще и национальным. В результате русскоговорящее население обладает большой вариативностью национальных акцентов и стилей произношения, которые весьма сложно учесть при создании дикторонезависимых систем распознавания речи, поскольку необходимо создавать очень большие базы данных обучения акустических моделей.

Для создания эффективных систем распознавания для английского языка существуют многочисленные речевые базы данных (обычной речи, телефонной и т.д.), в том числе и свободно доступные. Для русского языка такие базы данных пока только начинают создаваться и, как правило, являются закрытыми для общего пользования.

В следующих разделах будет предложено решение проблем автоматического распознавания русской речи, учитывающее указанные трудности и специфику русского языка.

2 Обучение системы распознавания русской речи

При создании дикторонезависимой системы распознавания речи возникает необходимость обучения акустико-лексических моделей этой системы.

На рис. 1 представлен процесс создания баз данных и словарей, необходимых для обучения дикторонезависимой системы распознавания речи. При распознавании речи используется словарь распознаваемых транскрибированных лексических единиц – слов или морфем (как в системе SIRIUS), поэтому при обучении системы распознавания речи и создании акустико-лексических баз данных необходимо использовать модули транскрибирования и морфемной сегментации.

Модуль транскрибирования осуществляет преобразование текстов предметной области в фонетическую транскрипцию. На вход модуля поступают: набор предложений, составляющих тексты; словарь слов из этих предложений, разбитый на морфемные единицы; словарь словоформ, полученных из базовых форм слов русского языка с отметкой ударного слога (слов), используемый фонетический алфавит и фонетические правила. Результатом работы модуля транскрибирования является транскрибированный словарь словоформ предметной области, разбитый на морфемы.

Продолжение таблицы 1

т	тост	т о! с т	м'	мята	м' а т а
т'	тень	т' е! н'	н	найти	н а й т' и!
д	дым	д ы! м	н'	нить	н' и! т'
д'	день	д' е! н'	л	луч	л у! ч
к	кот	к о! т	л'	любовь	л' у б о! ф'
к'	кит	к' и! т	р	краб	к р а! п
г	гусь	г у! с'	р'	резать	р' е! з а т'
г'	гибкий	г' и! п к' и й	й	июль	и й у! л'

Необходимой частью автоматических процедур словообразования и морфемной сегментации являются базы данных различных типов морфем, поэтому достаточно большая часть времени была уделена их созданию.

Построение базы данных морфем было осуществлено на основе печатных изданий. Большая часть корневых морфем (около 4 000) была взята из словаря морфем русского языка Кузнецовой [3], а различные фиксальные морфемы и флексии взяты из словаря «Русская грамматика» [4]. Кроме того, далее при создании целого ряда приложений словарь морфем постоянно пополнялся и сейчас составляет около 5 000 морфем.

На основе разработанных лексических баз данных можно строить любые приложения, лексика которых покрывается данным словарем. Поэтому разработанные в ходе исследования базы данных можно разбить на две группы: (1) словари и правила русского языка и (2) словарь, формируемый для конкретной предметной области (ПО). В ходе подготовки конкретного приложения также возможна модификация, дополнение основных словарей. Построение баз данных (в том числе, словаря морфем) конкретной предметной области начинается с анализа фраз ПО, формируется словарь ПО, далее производится разбиение слов на морфемы, при этом если находят новые морфемы, то они добавляются в исходный словарь морфем.

Разбиение слова на морфемы осуществляется путем подбора различных типов морфем с учетом правил следования морфем в одном слове. В табл. 2 возможные пары типов морфем отмечены знаком «+». При получении недопустимой пары «—» данная гипотеза разбиения слова на морфемы отбрасывается и поиск продолжается дальше, пока не обнаруживается конец слова «STOP».

Разработанные базы данных морфем использовались для создания морфемной модели языка, строящейся на основе статистики встречаемости различных пар морфем. Для первичной оценки работоспособности модуля создания модели языка были использованы доступные в Интернете текстовые корпуса. Из них был создан текст общим объемом около 50 Мб, который потом был предварительно обработан и все слова в нем были разбиты на морфемы. При этом сохранялась разметка на предложения и слова, а морфемы помечались, к какому типу они принадлежат. В результате анализа текста был получен словарь морфем размером около 5 000 морфем, разбитых на 5 типов, и получены вероятности встречаемости всех пар морфем.

Акустическое моделирование в нашей системе основано на Скрытых Марковских Моделях (СММ) [5]. Марковские модели довольно содержательны по своей математической структуре, поэтому они стали теоретическим фундаментом для различных областей исследований, а не только языка и речи. СММ-модель речевого сигнала позволяет решать не только задачи распознавания речи, но также

улучшать качество сигнала, загрязненного шумами и искажениями, моделировать источник речевого сигнала, оптимизировать структуру диалога и др.

Таблица 2 – Таблица согласования типов морфем в слове

Тип текущей морфемы	Тип следующей морфемы				
	Приставка	Корень	Интер-суффикс	Суффикс	Окончание
нет	+	+	—	—	—
Приставка	+	+	—	—	—
Корень	STOP	STOP	+	+	+
Интер-суффикс	+	+	—	—	—
Суффикс	—	—	—	+	+
Окончание	STOP	STOP	STOP	STOP	—

В СММ суть распознавания сводится к оценке правдоподобия принадлежности входной последовательности (наблюдаемой) к данной гипотетической модели. Проблема обучения акустических моделей очень важна для большинства приложений, так как именно во время обучения происходит, в соответствии с выбранным критерием оптимальности, подстройка значений параметров модели по данным наблюдений, в результате чего создается модель, наилучшим образом соответствующая реальному явлению. Каждая базовая распознаваемая единица (фонема, аллофон или трифон) представляется некоторым типом скрытой марковской модели, настройка параметров которой осуществляется по обучающему множеству речевых данных. Необходимо учитывать, что обучающее множество акустических данных для СММ должно быть достаточно велико, чтобы учитывать возможный спектр пользователей системы. В качестве акустических моделей мы использовали многокомпонентные непрерывные СММ с Гауссовской функцией распределения вероятностей появления векторов. Для параметризации сигнала использовались *mel*-частотные кепстральные коэффициенты с их первой и второй производной. Для выделения полезного речевого участка из шума нами был разработан и применен метод, основанный на вычислении спектральной энтропии сигнала [6]. Для определения границ речи используется критерий отличия значений энтропии для речевых сегментов и для фонового шума.

Фонемы использовались в виде трифонов (фонем в определенном фонетическом контексте). СММ трифонов имеют 3 значащих состояния (и 2 «пустых» состояния, предназначенных для объединения моделей трифонов в модели морфем). Процесс обучения СММ осуществлялся путем последовательного выполнения следующих процедур:

- инициализация СММ фонем;
- обучение СММ фонем на обучающих речевых БД (алгоритм Витерби);
- переход от моделей фонем к моделям трифонов и их инициализация;
- обучение СММ трифонов на обучающих речевых БД (алгоритм Витерби);
- объединение похожих моделей трифонов, имеющих небольшое количество обучающих данных;
- последовательное увеличение числа компонент гауссовых смесей моделей с одновременным обучением на речевых БД.

Для отработки методов автоматического обучения необходимо использование речевых баз данных, содержащих акустические файлы, а также их фонетические транскрипции, разметку акустического сигнала по фонемам, морфемам, словам и фразам. Для обучения СММ были заранее подготовлены обучающие тексты (отдельные предложения предметной области). На основе полученного материала была произведена вышеописанная процедура обучения акустических моделей (СММ трифонов).

Таким образом, описанный процесс формирования лексических баз данных для конкретной предметной области и обучение акустических моделей позволяет далее строить различные приложения, использующие распознавание речи. Разработанные методы и модули для распознавания русской речи, а также процесс обработки речевого сигнала подробно описаны в следующем разделе.

3 Механизм распознавания русской слитной речи

Подготовленные базы данных конкретной предметной области, а именно: транскрибированный словарь морфем ПО, морфемная модель языка ПО, список фонем ПО и набор акустических моделей фонем ПО далее используются в ходе распознавания речи. Однако если при подготовке баз данных мы анализировали фразы, разбивая их на фонемы, то теперь будет осуществляться обратный процесс: из наиболее вероятных цепочек фонем последовательно синтезируются морфемы, слова и фразы (рис. 2).

Речевой сигнал, поступающий с микрофона, в первую очередь проходит этап параметрического представления, где отрезаются начальные и конечные паузы в сигнале, а оставшийся участок кодируется в последовательность векторов признаков, которая уже следует в модуль распознавания фонем. При распознавании фонем (которые используются в форме трифонов) и формировании морфем используются методы Скрытого Марковского Моделирования и смесей Гауссовских распределений. В отличие от существующих аналогов, в нашей модели вместо слов используются морфемы. За счет этого на этапе распознавания лексических единиц было получено существенное увеличение в скорости. При этом по сравнению с распознаванием целых слов точность распознавания морфем несколько снизилась, но за счет последующих уровней обработки точность распознавания фраз практически не изменилась.

После распознавания фонем и подбора наиболее вероятных цепочек морфем получившийся набор гипотез далее используется для формирования цепочек слов. Синтез слов из различных типов морфем осуществляется по схеме, представленной на рис. 3.

В данной модели заданы начальное и конечное состояния, а в остальных узлах присутствуют все возможные типы морфем. Дугами обозначены возможные переходы. В будущем, возможно, эта модель будет вероятностной, а пока максимальное количество переходов из состояния в состояние заданы жестко. На этом этапе обработки на основе каждой поступившей гипотезы фразы, представленной в виде последовательности морфем, формируется еще несколько гипотез, представленных уже последовательностью гипотез слов.

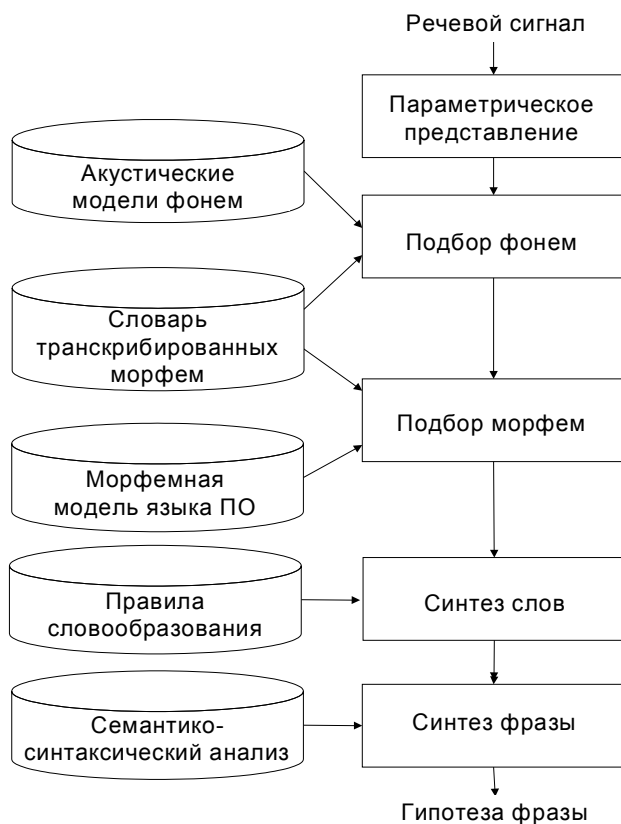


Рисунок 2 – Синтез фразы из речевого сигнала

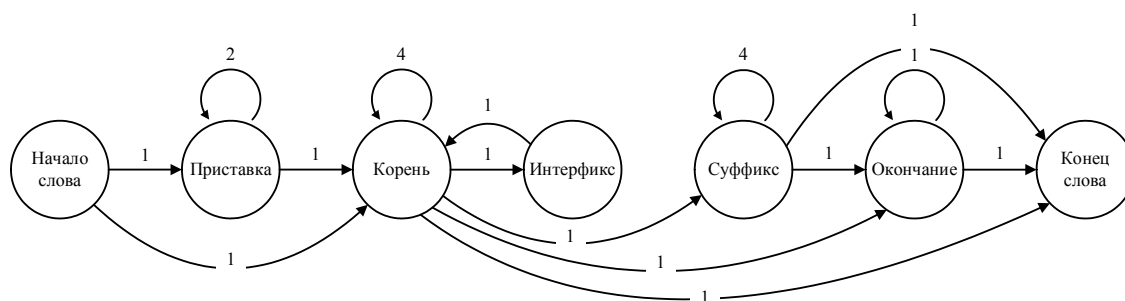


Рисунок 3 – Синтез слова из морфем

Последним этапом обработки является синтез фразы. Здесь за счет семантико-синтаксического анализа производится оценка поступающих гипотез и по интегральному критерию выбирается наилучшая. Данный модуль основан на ранее разработанной в группе речевой информатики СПИИРАН модели интегрального понимания речи и пока не подвергся особым изменениям [7].

Таким образом, в результате исследований созданы базы данных различных типов морфем русского языка, программные модули для автоматизации процессов словообразования и морфологического разбора, программный модуль для накопления статистики встречаемости морфем по текстовому материалу, а также другие программные средства, необходимые для создания системы распознавания русской речи с большим словарем [8].

4 Применение системы SIRIUS

Для отладки и тестирования система распознавания русской речи была внедрена в электронный каталог «Желтые страницы Санкт-Петербурга», который доступен по адресу www.yell.ru, с целью создания единой автоматизированной справочной системы для поиска адресов и телефонов организаций, необходимых пользователю, при помощи обычного телефона.

Создание автоматизированной справочной системы с голосовым доступом к ресурсам электронного каталога посредством телефона (автоматизированный call-центр) позволит использовать для ответов на звонки абонентов не операторов-людей, а систему автоматизированного распознавания речи с генерацией речевых ответов.

Размер словаря в данной задаче составил 1 850 слов, а при разбиении на морфемы размер морфемного словаря сократился до 1 360. Столь незначительное сокращение связано со спецификой задачи. Во-первых, такие части речи, как глагол, причастие, местоимение, имеющие особенно сложную структуру словообразования, употреблялись достаточно редко. Во-вторых, практически все существительные и прилагательные употреблялись только в именительном падеже во множественном числе (как это сделано в самом электронном каталоге). В то же время применение морфемного анализа при разборе произведения М.А. Булгакова «Мастер и Маргарита» дало уменьшение словаря почти в 8 раз (словарь слов – 22 984, словарь морфем – 2 920). Поэтому при создании стенографической системы реального времени с большим или средним словарем введение морфемного уровня будет иметь решающую роль в плане точности и скорости распознавания.

Для тестирования было предложено 635 фраз, записанных в офисных условиях. В эксперименте участвовало 5 дикторов. Записанные файлы были пропущены через модель целословного распознавания, а затем через морфемно-ориентированную модель распознавания с последующим словообразованием. Результаты экспериментов представлены в табл. 3. По сравнению с первой моделью, точность распознавания морфем несколько снизилась, но за счет последующих уровней обработки точность распознавания фраз практически не изменилась.

Также был проведен тест, направленный на сравнение скорости работы распознавателя, основанного на морфемном распознавании и целословного распознавателя. Результаты приведены в табл. 4. Общее количество тестовых фраз, содержащихся в тестовой базе данных составило 635 (состоящих из 2 574 слов). Из таблицы видно, что при использовании разработанной системы скорость возросла более чем в 1,7 раза, что при незначительном падении точности позволяет говорить о создании перспективной системы распознавания речи для больших словарей.

Таким образом, разработанный модуль распознавания слитной русской речи показал достаточно высокую точность и дикторнезависимость к носителям русского языка. Следующим этапом в разработке системы голосового доступа к электронному каталогу будет исследование диалогового режима с пользователем, а также настройка системы распознавания речи к условиям работы в телефонных линиях, создание речевых баз данных, записанных в процессе диалога пользователя с системой через телефонный канал.

Таблица 3 – Сравнение моделей распознавания по точности

Диктор	Целословное распознавание		Морфем-ориентированное распознавание	
	Точность распознавания слов	Точность распознавания фраз	Точность распознавания морфем	Точность распознавания фраз
1	96	94	82	92
2	93	90	78	90
3	95	93	81	92
4	91	91	79	91
5	95	93	82	91
Среднее	94,0	92,2	80,4	91,2

Таблица 4 – Сравнение моделей распознавания по скорости обработки

	Целословное распознавание	Морфем-ориентированное распознавание
Время, затраченное на тестовый набор	2993 с	1740 с
Среднее время на одну фразу	4,71 с	2,74 с
Среднее время на одно слово	1,16 с	0,67 с
Среднее время на одну морфему	–	0,47 с

5 Распознавание речи через телефонный канал

При внедрении системы распознавания речи для задачи распознавания телефонной речи возникает проблема ухудшения качества работы системы, связанная с различными характеристиками телефонных линий и телефонных аппаратов. Во-первых, полоса пропускания аналоговой телефонной линии ограничена полосой 0 – 4 000 Гц.

Чтобы уменьшить влияние шума, вносимого каналом передачи информации, были использованы два метода очистки сигнала: полосовой фильтр (для отсекаания некоторых низкочастотных шумов) и метод вычитания кепстрального среднего (Cepstral Mean Subtraction).

В настоящее время мы имеем результаты первых экспериментов по распознаванию телефонной русской речи. В качестве аппаратного обеспечения для соединения персонального компьютера с телефонной сетью общего пользования мы использовали внутренний голосовой факс-модем 3COM US Robotics и набор функций Telephone API для работы с телефонным устройством. Это наиболее простая версия соединения, и далее планируется использовать для соединения с

телефонной линией многоканальные платы Intel Dialogic для аналоговых телефонных линий.

Далее представлены результаты экспериментов по использованию CMS и полосового фильтра для распознавания телефонной речи. В ходе наших экспериментов наилучшие показатели продемонстрировал полосовой фильтр от 210 до 4 000 Гц. Рис. 4 показывает результаты по распознаванию речи в трех аналоговых телефонных линиях с различными техническими характеристиками (три столбца на диаграмме).

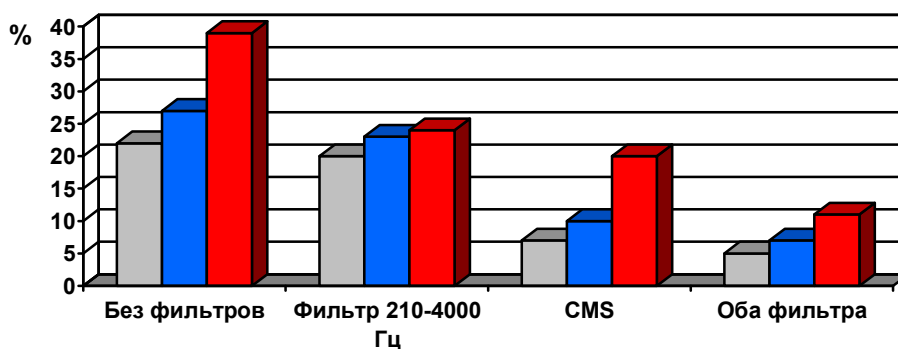


Рисунок 4 – Процент ошибок распознавания фраз при использовании различных методов фильтрации

Эта прикладная задача была связана с распознаванием произносимых через телефон русских цифр от 0 до 1000. Очевидно, что лучшие результаты показаны при использовании обоих методов фильтрации.

Таким образом, для решения задачи устойчивого распознавания речи в условиях телефонного разговора необходимо применение эффективных средств цифровой фильтрации и очистки сигнала от фонового шума. Для обучения и дальнейшего тестирования дикторонезависимой модели распознавания будет собрана база данных телефонной русской речи, созданная при использовании аналоговых телефонных линий.

Заключение

Исследование проблем автоматического понимания/распознавания речи является важным фундаментальным направлением. Эта проблема сейчас сдерживает развитие всевозможных систем взаимодействия человека с машиной. Представленная модель голосового интерфейса в первую очередь направлена на учет особенностей русской речи. Нами был разработан и опробован новый морфемный метод представления языка и речи. Он показал высокое качество и устойчивость работы на словаре до 2 000 слов конкретной предметной области. Разработаны базы данных различных типов морфем и получена статистика встречаемости пар морфем по текстовым корпусам. В результате такой обработки обеспечивается инвариантность к грамматическим отклонениям, а также увеличивается скорость распознавания русской речи и других языков со сложным механизмом словообразования (в частности, славянских языков).

Последующие работы направлены на увеличение размера распознаваемого словаря, а также настройку системы распознавания русской речи SIRIUS к работе с телефонным каналом. При внедрении системы распознавания речи в телекоммуникационные приложения будут учтены проблемы, связанные со спецификой телефонных линий и различиями характеристик телефонных аппаратов. Накопленные речевые и лексические базы данных будут использованы для дальнейшего изучения механизма понимания речи и создания эффективных средств человеко-машинного взаимодействия.

Литература

1. Ronzhin A.L., Karpov A.A. Implementation of morphemic analysis for Russian speech recognition / Proc. of 9th International Conf. SPECOM'2004. – Russia. – St. Petersburg, Publishing house «Anatolya». – 2004. – P. 291-296.
2. Грамматический словарь русского языка / Сост. А.А. Зализняк. – М., 1977.
3. Словарь морфем русского языка / Сост. А.И. Кузнецова, Т.Ф. Ефремова. – М.: Рус. яз., 1986. – 1136 с.
4. Русская грамматика / АН СССР. Ин-т рус. яз. – М.: Наука, 1980.
5. Young S., et al. The HTK Book (v3.0). – Cambridge Univ. Engineering Department. – 2000.
6. Карпов А.А. Робастный метод определения границ речи на основе спектральной энтропии / Искусственный интеллект. – 2004. – № 4. – С. 607-613.
7. Методы понимания речи и текста / Ю.А. Косарев, И.В. Ли, А.Л. Ронжин, J. Savage / Под ред. Р.М. Юсупова // Труды СПИИРАН. – Т. 2, вып. 1. – СПб.: «Анатолия», 2002. – С. 157-195.
8. Ronzhin A.L., Karpov A.A. Large Vocabulary Automatic Speech Recognition for Russian Language / Proc. of Second Baltic Conf. on Human Language Technologies. – Tallinn (Estonia). – 2005. – P. 329-334.

А.Л. Ронжин, О.А. Карпов, I.V. Ли

Система автоматичного розпізнавання російського мовлення SIRIUS

У статті представлена розроблена в групі мовленнєвої інформатики СПИРАН система розпізнавання стягненого російського мовлення SIRIUS. Особливістю даної системи є наявність у ній морфемного рівня представлення мови і мовлення, що дозволяє значно скоротити розмір лексичного розпізнаваного словника і швидкість обробки. Описано процес упровадження системи розпізнавання російського мовлення в галузі інфотелекомунікацій для голосового доступу до інтернет-версії електронного каталогу «Жовті сторінки Санкт-Петербурга» з метою створення автоматизованого call-центра для відповідей на дзвінки абонентів. У статті представлено результати тестування системи як при роботі з мовленням, записаним в офісних умовах, так і в умовах телефонної розмови.

O.L. Ronzhin, A.A. Karpov, I.V. Lee

Automatic System for Russian Speech Recognition SIRIUS

In the paper the developed in Speech Informatics Group of SPIIRAS speech recognition system SIRIUS is presented. The peculiarity of this system is presence in this system the morphemic level of speech and language representation, that allows essential reducing of the size of lexical vocabulary and speech of processing. The process of introduction of ASR into infotelecommunications area for voice access to electronic catalogue «Yellow Pages of St. Petersburg», aimed for creation of automatic call-center for answering on users' calls, is presented. In the paper the results of testing the system for work with clear speech and telephone speech is presented.

Статья поступила в редакцию 11.07.2005.