

# Экспериментальная оценка качества распознавания разговорной русской речи

*Кибкало А.А., кандидат физико-математических наук,*

*Данилов А.Г., ОАО «ИТЦ «Система-Саров»,*

*Мышкин А.С., ОАО «ИТЦ «Система-Саров»,*

*Сысоев В.Н., ОАО «ИТЦ «Система-Саров»,*

*Ромашкин Ю.Н., кандидат технических наук*

В данной статье излагаются основные проблемы и результаты решения задачи автоматического распознавания речи, полученные ведущими иностранными компаниями. Приводятся результаты испытаний опытного образца аппаратно-программного комплекса автоматического преобразования русской разговорной речи в текст, разработанного ОАО «ИТЦ «Система-Саров».

• автоматическое распознавание речи • скрытые марковские модели • русский язык • ошибка распознавания слов.

Main problems of automatic speech recognition task and results of leading developer groups are presented. The prototype of the hardware-software complex for Russian spontaneous speech-to-text conversion designed by ITC «Sistema-Sarov» company is described. Results of test evaluations of the complex are presented.

• automatic speech recognition • hidden markov model • russian language • word error rate.

## Введение

Первые разработки в области распознавания слитной речи были ориентированы на речь, записанную в студии в благоприятной акустической обстановке, с использованием высококачественных микрофонов. При этом дикторы зачитывали текст, разбитый на отдельные грамматически правильные фразы. Такую речь обычно называют прочитанной. Классической задачей распознавания прочитанной речи стала задача Wall Street Journal (WSJ) [1], характеризуемая следующими основными параметрами:

- большой словарь, содержащий 20 тысяч слов (такие декодеры речи, как правило, применимы и для задач со словарём в 100–150 тысяч слов);
- статистические грамматики для большого словаря (триграмные и категориальные языковые модели);



- постановка проблема распознавания слов, отсутствующих в словаре (OOV слов);
- различные методы адаптации акустических и языковых моделей;
- различные оптимизационные техники (трифонное дерево словаря, предвычисление языковых и акустических вероятностей, быстрое вычисление гауссовых смесей и т.д.).

В реальной жизни люди редко пользуются речью, напоминающей прочитанную. Первым серьезным шагом в распознавании естественной речи стала задача Air Travel Information Systems (ATIS) [2, 3], в которой требовалось распознавать голосовые запросы к базе данных информации об авиасообщении. В задаче ATIS появились новые моменты:

- более естественный стиль речи;
- речь, записанная в типовом офисном окружении;
- элементы понимания речи (принятие решения по распознанным ключевым словам) и соответствующая метрика оценки;
- отдельное моделирование неключевых слов (filler model);
- разработанный сценарий речевого взаимодействия человека и компьютера (он стал типичным для современных call-центров, речевые сервисы которых занимают большую часть рынка речевых технологий).

Задача транскрипции теленовостей Broadcasting News (BN) [4] фактически стала переходной к задаче распознавания естественной разговорной речи. Основные особенности этой задачи и методов её решения:

- речь профессиональных дикторов и корреспондентов несколько формализована, речь их собеседников ближе к разговорной;
- сплошной поток речи вместо декодирования отдельных фраз, требующий сегментации речи по фразам;
- речь нескольких дикторов, требующая распознавания диктора и сегментации речи по дикторам, редкая смена дикторов;
- необходимость отделения речи от неречевых фрагментов (музыки, шума и т.д.);
- чередование широкополосной и телефонной речи;
- задача BN не является искусственной и близка таким реальным задачам, как, например, автоматическое протоколирование заседаний.

Лучшие результаты, достигнутые в задаче BN, 10–15% ошибок при распознавании слов в режиме реального времени. Способность получить хорошие результаты в этой задаче стала фактически критерием состоятельности команды разработчиков систем автоматического распознавания речи (далее APP). Вслед за задачей транскрипции теленовостей на английском языке появились её клоны для других языков (японского испанского, китайского и др.), за исключением, к сожалению, русского [5].

Спонтанной или разговорной обычно называют речь, используемую людьми в свободной обстановке: телефонных разговорах, неформальных обсуждениях и т.д. Акустику, лексику и грамматику такой речи моделировать значительно сложнее. Но наибольший интерес с практической точки зрения представляет распознавание именно спонтанной речи.

В 1993 году появился первый релиз корпуса спонтанной речи — Switboard (SB) [6]. Это был первый корпус, содержащий телефонные диалоги (более 240 часов речи 541 диктора). Корпус сопровождался многоуровневой транскрипцией (без фонетики) и предназначался для работ по различным на-

правлениям речевых технологий. Затем появился корпус CALLHOME (CH) [7]. Задача распознавания спонтанной речи для японского языка была сформулирована в [8]. Работа над ней началась с создания большого корпуса (800 часов речи, 7 миллионов слов в текстовом корпусе). Речь в этом корпусе содержала все типичные для естественной разговорной речи явления: паузы, слова-заполнители, нефонетические звуки, повторы, не полностью произнесённые слова и фразы, исправления слов и фраз, слабо артикулированные звуки, различия между звуками [9].

Декодирование такой речи столкнулось с серьёзными трудностями. Первые системы делали 40–60% ошибок, лучшие результаты — 20–25% для SB и 30–35% для CH [10]. Позже появились подобные корпуса и для других распространённых языков (опять-таки кроме русского). На этих корпусах были выработаны основные подходы к распознаванию спонтанной речи.

Результаты испытаний систем APP на задачах SB и CH показали, что, хотя и был достигнут определённый уровень в технологии распознавания спонтанной речи, для дальнейшего продвижения необходима корректировка традиционного представления системы APP как совокупности акустической, лексической и языковой моделей. Для успешного понимания речи необходимо распознавание не только произнесённых слов, но и других событий и явлений (имеющих как лингвистическую, так и нелингвистическую природу). К первым, помимо выше перечисленных, относятся: границы фраз и частей фраз, соответствующие знакам пунктуации; типы фраз (вопросительный, утвердительный и т.д.); имена собственные; аббревиатуры и т.д.

Нелингвистические события и факторы включают: смену диктора; одновременную речь нескольких говорящих; характеристики диктора (пол, возраст и т. д.); нефонетические звуки (кашель, смех, стук двери, скрип и т.д.).

Для представления этих событий и явлений используются так называемые метаданные, обычно представляемые в похожем на XML формате [11]. При таком подходе задача распознавания спонтанной речи распадается на две основные подзадачи — перевод речи в текст и аннотация метаданных.

Новый подход к разработке и оценке систем распознавания спонтанной речи был положен в основу программы EARS (Effective Affordable Reusable Speech) [12] и серии тестовых задач проекта Rich Transcription (2002–2009) [13]. В первых задачах проекта RT использовались данные корпусов BN и SB, в последних — корпуса, подготовленные участниками испытаний (лекции с обсуждением, интерактивные презентации, разговоры в перерывах). Основные характеристики задач:

- наличие границы слов и нефонетических звуков в разметке обучающей части корпуса;
- небольшое количество дикторов (3–6) в каждом сеансе;
- участие «ненативных» дикторов;
- запись с разных микрофонов.

При аннотации метаданных на разных этапах проекта RT оценивались наличие речи, диаризация дикторов, определение границ фраз, распознавание фонетических и нефонетических явлений и звуков. В последних испытаниях проводилась также интегрированная оценка точности распознавания и диаризации дикторов (Speaker Attributed STT) [14].

## Распознавание русской речи

В настоящее время уровень технологий распознавания русской речи в целом уступает аналогичным показателям для других распространённых языков [15]. Это обусловлено комплексом причин:

- естественной сложностью русского языка для автоматического распознавания;



- отсутствием или труднодоступностью русских языковых ресурсов;
- сравнительно малым объемом рынка отечественных речевых технологий.

Из свойств русского языка, усложняющих реализацию систем распознавания слитной речи, по отношению к языкам, для которых велись активные работы в этой области, наиболее важными являются высокая флективность (большое количество словоформ) русского языка и произвольный порядок слов в предложении.

Высокая флективность русского языка вызвана изменением русских слов по падежам, родам, числам и временам. Так, задача распознавания речи на английском или китайском языках, включающая 100 тысяч разнокоренных слов, для русского языка эквивалента не менее двум миллионам словоформ. Большое количество словоформ предъявляет дополнительные требования к точности и скорости работы алгоритмов поиска, используемых при распознавании русской речи.

Произвольный порядок слов в русском предложении значительно снижает эффективность традиционных языковых моделей, используемых при распознавании слитной речи. Так, основной показатель эффективности традиционно используемой  $n$ -граммной языковой модели — её неопределённость (среднее количество альтернатив для следующего слова) в русском языке в несколько раз больше аналогичного показателя для английского или китайского языков. Для текстовых корпусов английского и русского языков одинакового размера и словарей, обеспечивающих 0,5% OOV слов, неопределённость триграммной модели для английского языка равна 224, а для русского — 714 [16]. Поэтому требуется разработка эффективной модели русского языка с высокой неопределённостью.

Ряд работ был посвящён статистическим языковым моделям для языков с повышенной флективностью. В первой такой работе [16] неопределённость уменьшалась до значений порядка 400 при использовании категорных моделей, но реальное распознавание русской речи не выполнялось. Несколько лучшие характеристики моделей были получены для интерполяции триграммной модели и модели, основанной на случайных лесах [17], двухпроходной языковой модели с расширением словаря на втором проходе [18]. Языковая модель с малой неопределённостью была предложена в [19].

Серьёзным фактором, затрудняющим разработки в области распознавания русской речи, являются проблемы, связанные с русскими языковыми ресурсами. Хотя в России и был записан ряд аудиокорпусов русского языка, ситуация с доступом к ресурсам остаётся непростой. Корпус ISABASE содержит временную разметку по звукам и может использоваться на начальной стадии построения акустических моделей. Речь дикторов не совсем естественна, слова произносятся подчёркнуто и с большими паузами. Корпус RuSpeech, записанный компанией Cognitive Technologies — первый русский корпус хорошего качества. Корпус SPEECHRU (компания Одитек) содержит 4 канала речи с различным уровнем шумов и записан согласно стандартам проекта SpeechDat. Наиболее значительные русские языковые ресурсы находятся в каталоге ELDA [20], но их стоимость достаточно высока. Русские ресурсы есть также в репозиториях SpeechOcean [21] и LDC [22]. Среди русских ресурсов практически отсутствуют текстовые корпуса, ориентированные на речевые технологии.

К языковым ресурсам можно также отнести стандартные тестовые задачи. Они включают формулировку задачи, языковые ресурсы (аудиокорпус и текстовый корпус) и методику оценки. Сравнительные испытания на стан-

дартных тестовых задачах позволяют объективно оценить достижения разработчиков систем распознавания речи. Подготовка таких задач является одной из важнейших сторон деятельности речевого подразделения американского института NIST. К сожалению, не существует каких-либо стандартных тестовых задач для русских речевых технологий.

Системы автоматического распознавания русской речи пока находятся стадии экспериментов [15]. Причём, либо представлены результаты для медленно работающих (0,8–0,9 реального времени) систем с весьма небольшим словарём (1850 слов) [23], либо приведены данные, не позволяющие объективно оценить характеристики систем [24].

В 2010–2012 гг. ОАО «ИТЦ «Система-Саров» по государственным заказам выполнило НИ-ОКР [25–27], целью которых была разработка и испытания опытного образца аппаратно-программного комплекса (АПК) автоматического преобразования разговорной русской речи в текст.

Важной их составной частью было создание русского телефонного корпуса спонтанной речи объёмом 200 часов. Корпус был записан и размечен компанией «Одитек» (г. Санкт-Петербург) и впоследствии использован в ОКР для построения акустических моделей. При создании и использовании этого корпуса возникли следующие проблемы:

- общий объём корпуса существенно уступал зарубежным аналогам;
- отсутствовал текстовый корпус большого размера (по Фуруи — 7 млн слов), что затруднило построение адекватных языковых моделей;
- объём и сложность работ по разметке корпуса спонтанной речи оказались существенно больше, чем для аналогичных корпусов прочитанной речи, в связи с чем выросло и время создания корпуса.

## Архитектура разработанного АПК

В целом задача состояла в разработке и реализации речевых технологий для семантического анализа разговорной русской речи, а также создании среды, позволяющей эффективно применять эти технологии. В качестве базового программного обеспечения был выбран пакет программ SDT [28]. Рабочая среда АПК состоит из русского речевого MRSP-двигателя, компоненты управления заданиями и базы данных. Структурные схемы аппаратной части АПК и программного обеспечения приведены на рис.1 и 2. В состав АПК включены два сервера обработки HP Proliant DL360 G7 X5675 / 3.0 ГГц 2x6 ядер / 16 Гб ОЗУ.

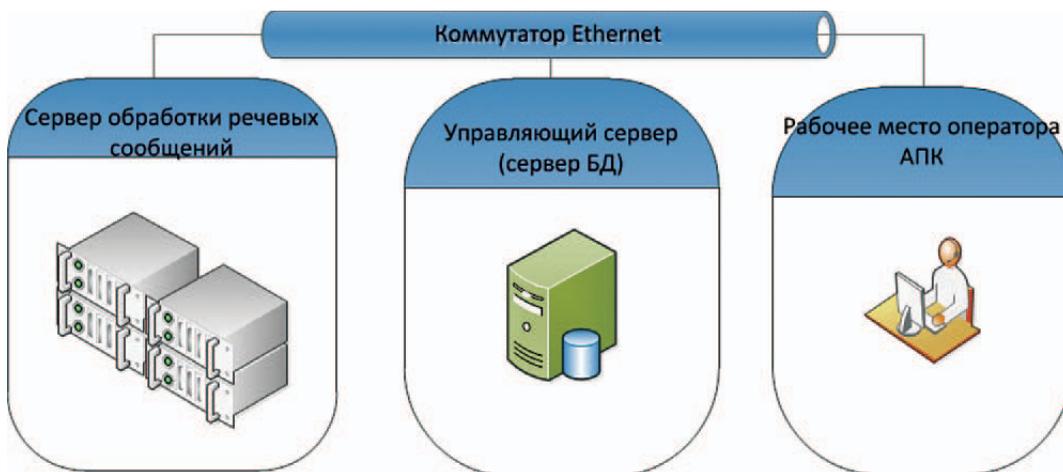


Рис. 1. Структурная схема аппаратной части АПК

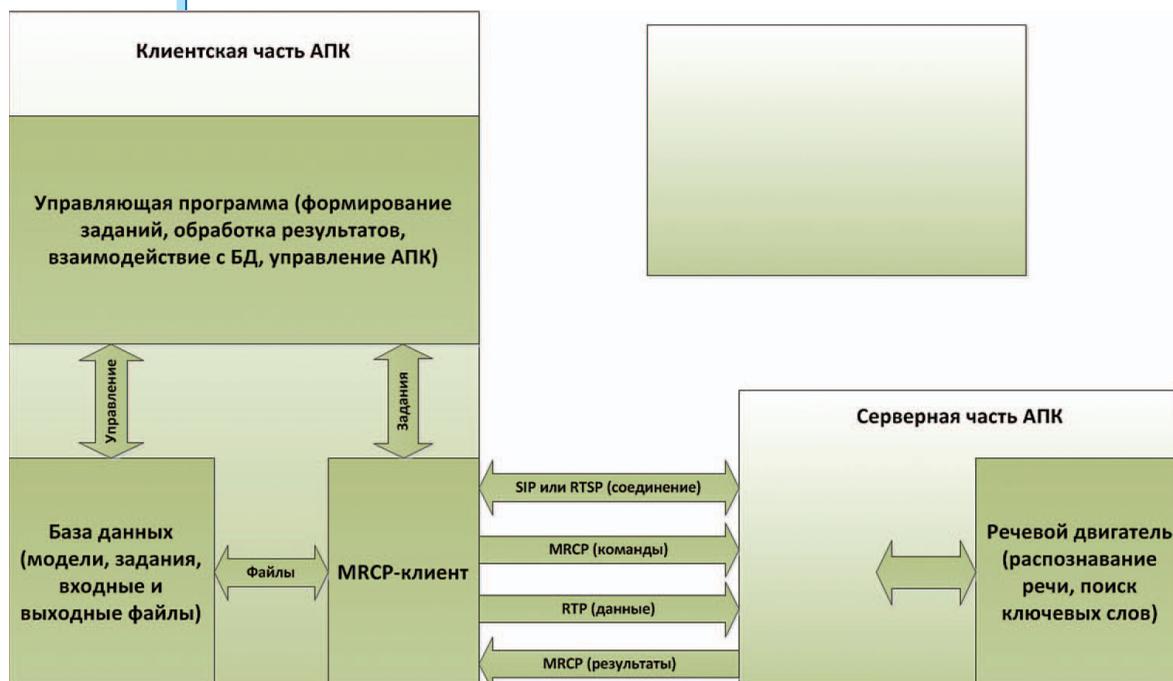


Рис. 2. Архитектура программного обеспечения АПК

Для перевода речевого сообщения на русском языке в текстовую форму используются акустические, лексические и языковые модели. Акустическая модель использует непрерывные скрытые марковские модели с многокомпонентными гауссовыми распределениями вероятностей. Лексическая модель (фонетический словарь) представляет собой список допустимых слов и их транскрипций в алфавите SAMPA. Для моделирования языка используется конечная грамматика или триграмная статистическая модель. При распознавании речи аудиоданные предварительно разбиваются на фразы, и каждая фраза распознаётся независимо.

Протокол MRCP [29, 30] является открытым стандартом для речевых сервисов в call-центрах. Его применение позволяет использовать любой речевой двигатель, поддерживающий этот протокол, а также облегчить масштабирование комплекса. В опытном образце АПК, разработанном ОАО «ИТЦ «Система-Саров», используется русский речевой двигатель на базе пакета SDT. Поддержка протокола MRCP реализована с использованием открытых пакетов Sophia-SIP [31] и Unistack MRCP [32].

Компонента управления моделями обеспечивает управление акустическими, лексическими и языковыми моделями. Она позволяет импортировать акустические модели в среду АПК, а также адаптировать эти модели для конкретного диктора или группы дикторов. Для построения лексической модели оператор формирует список слов, транскрипции которых строятся автоматически с использованием универсального фонетического словаря и автоматического транскриптора. Оператор при необходимости может вручную исправить транскрипции отдельных слов. Языковая модель строится для выбранной акустической модели и текстовых файлов. При этом в процессе построения модели формируется фонетический словарь. Для построения языковой модели используются пакеты CMU Language Model Toolkit [33] и SDT [28].

База данных АПК содержит архив заданий на обработку, входные и выходные файлы заданий, а также используемые акустические, лексические и языковые модели. База данных обеспечивает также экспорт содержимого в стандартном формате, поддерживаемом большинством СУБД.

Пользовательский интерфейс программного обеспечения АПК включает функции аутентификации пользователя и доступа к компонентам управления заданиями и моделями. Он основан на web-интерфейсе, использующем платформу Drupal. Основные рабочие окна интерфейса: аутентификация, запуск заданий, список задач, работа с моделями. Пример окна списка задач показан на рис.3.

Первая страница | Запуск заданий | Список задач | Выход

Первая страница

Временной период: с 01.01.2012 по 30.07.2012

Опции выбора задач:  
Тип:  Статус:  Языковая модель:  Акустическая модель:

Выбрано задач: 0

Показывать по 10 записей

#	Тема	Тип	Статус	Язык.модель	Акуст.модель	Дата
98	Радио	распознавание	завершено	общая лексика	стандартная	2012.07.27 16:35:01
95	Радио	распознавание	завершено	общая лексика	стандартная	2012.07.26 11:35:22
94	Радио	распознавание	ошибка	общая лексика	стандартная	2012.07.26 11:31:55
93	Газеты	распознавание	в обработке	медицина		2012.07.23 14:31:53
92	Радио	диаризация	завершено	общая лексика	стандартная	2012.07.10 17:04:28
91	Радио	поиск ключевых слов	в обработке	общая лексика	стандартная	2012.07.10 17:04:17
90	Радио	распознавание	в обработке	общая лексика	стандартная	2012.07.10 17:04:01
89	Радио	диаризация	завершено	общая лексика	стандартная	2012.07.10 17:03:32
88	Радио	поиск ключевых слов	в обработке	общая лексика	стандартная	2012.07.10 17:03:24
87	Радио	распознавание	в обработке	общая лексика	стандартная	2012.07.10 17:03:08

Показано с 1 по 10 из 96 записей

в начало << 1 2 3 4 5 >> в конец

Работать с группой | Выделить текущие задачи | Снять выделение

Рис. 3. Окно «Список задач»

## Условия и результаты экспериментов

Для построения акустических моделей в испытаниях АПК использовались телефонные части аудиокорпусов [27, 35] общим объёмом примерно 200 часов речи. Акустическая модель включала около 6 тыс. состояний и 72 тыс. гауссианов.

Качество работы АПК оценивалось по следующим показателям:

- вероятность правильного распознавания слов в слитной речи (с учётом правильного согласования их окончаний, времени, рода и числа):

$$\hat{P}_D = \frac{N_D}{N},$$

где  $N_D$  — количество слов, правильно автоматически распознанных и согласованных в предложениях;

$N$  — общее количество слов (без предлогов) в эталонных текстах звуковых файлов;

— вероятность ошибочного распознавания слов:

$$\hat{P}_E = \frac{N_{ins} + N_{del} + N_{sub}}{N},$$

где  $N_{ins}$  — количество ошибок типа «вставка лишнего слова»;

$N_{del}$  — количество ошибок типа «пропуск слова»;

$N_{sub}$  — количество ошибок типа «замена слова на другое»;

— коэффициент ускорения:

$$\hat{K} = \frac{T_\phi}{T_{обр.}}$$

где  $T_\phi$  — суммарная длительность звуковых файлов, подлежащих обработке,

$T_{обр.}$  — время их обработки на АПК.

Тестовые звуковые файлы были записаны в режиме «моно» и содержали диалоги длительностью не менее 1 минуты произвольных абонентов в каналах телефонной связи общего пользования. Для каждого диалога имелись их текстовые расшифровки. Результаты испытаний представлены в таблице. Следует отметить, что как вероятность правильного распознавания речи, так значение коэффициента ускорения АПК зависят от степени соответствия содержания диалогов используемой модели русского языка (при плохом соответствии получили низкие значения вероятности и коэффициента ускорения, при хорошем — более высокие).

Таблица

Размер фонетического словаря	$\hat{P}_D, \%$	$\hat{P}_E, \%$	$\hat{K}$
85 тыс.	40–70	34–68	9–28

## Заключение

Высокая флективность русской речи, произвольный порядок слов в предложениях и небольшой объём доступных языковых ресурсов определяют особую сложность задачи её автоматического распознавания по сравнению с другими наиболее распространёнными языками. Тем не менее, соответствующие разработки проводятся разными российскими организациями, в том числе и ОАО «ИТЦ «Система-Саров».

Результаты испытаний опытного образца АПК автоматического распознавания разговорной русской речи в каналах телефонной связи общего пользования, разработанного ОАО «ИТЦ «Система-Саров», показали, что пока до-

стигается вероятность правильного распознавания от 40 до 70%, и она зависит от степени соответствия содержания речи используемой модели языка.

Для дальнейшего повышения показателей точности распознавания необходимо расширение корпуса используемых ресурсов (в первую очередь, текстового корпуса) и реализация более эффективной модели русского языка, учитывающей её высокую неопределённость.

## Литература

1. Pallett D.S., Fiscus J.G., Fisher W.M. и др. Benchmark Tests for the ARPA Spoken Language Program, 1994.
2. Price P.J. Evaluation of Spoken Language Systems the ATIS Domain // **Proc. of the Workshop on Speech and Natural Language**, 1990. P. 91–95.
3. Palett D.S., Dahlgren N.L., Fiscus J.G. и др. ATIS Benchmark Test Results / DARPA, 1992. [http://www.itl.nist.gov/iad/mig/tests/rt/ASRhistory/pdf/darpa92\\_atis.pdf](http://www.itl.nist.gov/iad/mig/tests/rt/ASRhistory/pdf/darpa92_atis.pdf).
4. Graff D. The 1996 Broadcast News Speech and Language-Model Corpus // **Proc. of the 1997 DARPA Speech Recognition Workshop**. P. 11–14.
5. Palett D.S., Fiscus J.G., Garafolo J.S. и др. Broadcast News Benchmark Test Results English and Non-English Word Error Rate Performance Measures, 1998. [http://www.itl.nist.gov/iad/mig/tests/rt/ASRhistory/pdf/bnews98\\_tests.pdf](http://www.itl.nist.gov/iad/mig/tests/rt/ASRhistory/pdf/bnews98_tests.pdf).
6. Wheatley B., Doddington G., Hemphill C. и др. A User's Manual // Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA, 1995. [http://www ldc.upenn.edu/Catalog/readme\\_files/switchboard.readme.html](http://www ldc.upenn.edu/Catalog/readme_files/switchboard.readme.html).
7. Canavan A., Graff D., Zipperlen G. CALLHOME American English Speech // Linguistic Data Consortium. <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC97S42>.
8. Maekawa K., Koiso H., Furui S. и др. Spontaneous Speech Corpus of Japanese // Proc. of LREC Conference, Athens, Greece, 2000. P. 947–952.
9. Furui S., Nakamura M., Ichiba T. и др. Why Is the Recognition of Spontaneous Speech so Hard? // Proc. of Text, speech and dialogue conference, Karlovy Vary, Czech Republic, 2005. P. 9–22.
10. Chase L.L. A Review of the American Switchboard and Callhome Speech Recognition Evaluation Programs // Proc. of LREC Conference, Granada, Spain, 1998. P. 789–794.
11. Guidelines for RT-03 Transcription — Version 2.2 / S. Strassel. Linguistic Data Consortium, University of Pennsylvania, 2003. [http://projects ldc.upenn.edu/Transcription/rt-03/RT\\_Transcription\\_V2.2.pdf/](http://projects ldc.upenn.edu/Transcription/rt-03/RT_Transcription_V2.2.pdf/)
12. Wayne C. Human Language Technology TIDES, EARS, Babylon // DARPA Tech Symposium, Anaheim, CA, July 2002.
13. Garafolo J.S., Fiscus J.G., Martin A. и др. NIST Rich Transcription 2002 Evaluation: A Preview // Proc. of LREC Conference, Las Palmas, Spain, May 2002. P. 1928–1932.
14. The 2009 (RT-09) Rich Transcription Meeting Recognition Evaluation Plan / Linguistic Data Consortium. [www.itl.nist.gov/iad/mig/tests/rt/2009/docs/rt09-meeting-eval-plan-v2.pdf](http://www.itl.nist.gov/iad/mig/tests/rt/2009/docs/rt09-meeting-eval-plan-v2.pdf)
15. Ронжин А.Л., Ли И.В. Автоматическое распознавание русской речи // Вестник Российской Академии Наук, 2007. Т. 77, № 2. С. 133–138.
16. Whittaker E.W.D. Statistical Language Modelling for Automatic Speech Recognition of Russian and English: Ph.D. Thesis, University of Cambridge, Cambridge, Great Britain, 2000.



*Oparin I.* Language Models for Automatic Speech Recognition of Inflectional Languages: Ph.D. Thesis, Plzen, 2008.

17. *Podvesky P., Machek P.* Speech Recognition of Czech: inclusion of rare words helps // Proc. of ACL Student Research Workshop, Ann Arbour, MI, 2005.

18. *Холоденко А.Б.* О построении статистических языковых моделей для систем распознавания русской речи // Интеллектуальные системы, 2002. Т. 6, вып. 1–4. С. 381–394.

19. Evaluations and Language resources Distribution Agency <http://www.elda.org/rubrique1.html>

20. SpeechOcean (Beijing Haitian Ruisheng Science Technology Ltd.). <http://www.speechocean.com/index.html>.

21. Linguistic Data Consortium <http://www ldc.upenn.edu/>

22. *Ронжин А.Л., Карпов А.А., Ли И.В.* Система автоматического распознавания русской речи SIRIUS // Искусственный интеллект, № 3, 2005.

23. *Tatarnikova M., Tangel I., Oparin I.* и др. Building Acoustic Models for a Large Vocabulary Continuous Speech Recognizer for Russian // Proc. of SPECOM Conference, St. Petersburg, June 2006. P. 83–87.

24. Техническое задание на НИР «Разработка и исследование алгоритмов автоматического распознавания разговорной русской речи» («Ш-2010–08–2.3»). ФГКУ «В/ч 35533», 2010.

25. Техническое задание на НИР «Создание аннотированного корпуса спонтанной русской речи применительно к проводным линиям (аналоговым и цифровым) телефонной связи общего пользования» (в рамках НИР «Ш-2010–08–2.3»). ОАО «ИТЦ «Система-Саров», 2010.

26. Техническое задание на ОКР «Разработка аппаратно-программного комплекса автоматического распознавания русской речи» («Ш-2011–08–2.4»). ФГКУ «В/ч 35533», 2011.

27. *Баранников В.А., Кибкало А.А.* Пакет программ построения систем распознавания речи // Труды III Всероссийской конференции «Теория и практика речевых исследований» АРСО. Москва, МГУ им. М.В. Ломоносова, 2003.

28. A Media Resource Control Protocol (MRCP) / IETF RFC 4463 Document, April, 2006. <http://tools.ietf.org/html/rfc4463>

29. Media Resource Control Protocol Version 2 (MRCP v2) / IETF Internet Draft, March, 2011. <http://tools.ietf.org/html/draft-ietf-speechsc-mrcpv2-24>

30. Sofia-SIP Library. <http://sofia-sip.sourceforge.net>

31. Open Source MRCP Project (UniMRCP). <http://www.unimrcp.org>

32. The CMU Statistical Language Modeling Toolkit. [http://www.speech.cs.cmu.edu/SLM\\_info.html](http://www.speech.cs.cmu.edu/SLM_info.html)

33. Open Source Content Management Platform Drupal. <http://drupal.org>

34. Сайт компании Одитек <http://auditech.ru/>

35. *Kibkalo A.A., Turovets A.A.* Basic Pronouncing Dictionary for Russian Speech Decoding // Proc. of SPECOM Conference, Moscow, 2003.

36. *Kibkalo A.A., Rogozhkin I.G.* The Russian Speaker Independent Consecutive Speech Decoder // Proc. of SPECOM Conference, Moscow, 2003.

37. *Demuyck K.* Extracting, Modelling and Combining Information in Speech Recognition: Ph.D. Thesis, Katholieke Universiteit Leuven, Belgium, 2001.

**Сведения об авторах:**

***Кибалко Александр Алексеевич —***

кандидат физико-математических наук, ОАО «ИТЦ "Система-Саров"», начальник отдела, окончил мех-мат МГУ, 1978, распознавание речи, распознавание образов.

E-mail: alexander.kibkalo@gmail.com

***Данилов Александр Геннадьевич — Г., Мышкин А.С., Сысоев В.Н. —***

ОАО «ИТЦ "Система-Саров"», программист, окончил мех-мат МГУ, 2004, распознавание речи, теория автоматов.

alexander.danilov@sarov-itc.ru

***Мышкин Александр Сергеевич — Сысоев В.Н. —***

ОАО «ИТЦ "Система-Саров"», главный программист, окончил МИФИ, 1993, распознавание речи, речевые кодеки.

alexander.myshkin@sarov-itc.ru

***Сысоев Владимир Николаевич —***

ОАО «ИТЦ "Система-Саров"», главный программист, окончил Балтийский ГТУ, 1997, распознавание речи, автоматизированные системы.

vladimir.sysoev@sarov-itc.ru

***Ромашкин Юрий Николаевич —***

кандидат технических наук, окончил Московский инженерно-физический институт, факультет «Автоматика и электроника». Область научных интересов: цифровая обработка речевых сигналов, фильтрация речи на фоне помех, автоматическое распознавание речи и языка, идентификация говорящего по голосу, низкоскоростное кодирование речи, оценка качества трактов речевой связи.

E-mail: romayn@yandex.ru